DOCUMENT RESUME

ED 429 130 TM 029 699

TITLE	Learning from Consumer-Oriented Review Efforts To Guide the Development of a System of Expert Panels To Identify and Share Promising and Exemplary Products and Programs. Working Papers. Volumes One and Two.
INSTITUTION	Office of Educational Research and Improvement (ED), Washington, DC. Office of Reform Assistance and Dissemination.
REPORT NO	ORAD-1999-1203-I; ORAD-1999-1203-II
PUB DATE	1999-03-00
NOTE	651p.; Volume 1 contains eleven papers, identified as A-K. Volume 2 contains 16 papers, identified as L-AA.
PUB TYPE	Collected Works - General (020)
EDRS PRICE	MF03/PC27 Plus Postage.
DESCRIPTORS	Consumer Economics; *Educational Improvement; Elementary Secondary Education; Evaluation Methods; Identification; Information Dissemination; Program Development; Program Effectiveness; Program Evaluation; Research Methodology; *Standards; *User Needs (Information)
IDENTIFIERS	*Experts; Office of Educational Research and Improvement

ABSTRACT

These volumes contain working papers related to the development of the Office of Educational Research and Improvement (OERI) Facilitated System of Expert Panels. Introductory papers by Susan Klein, Michael Scriven, Sharon Bobbitt and Susan Klein, and Michael Scriven and Lois-ellin Datta (printed in both volumes) set the stage for the discussions of expert panels. Section two contains commissioned papers on review activities designed to find the best practices and procedures and the third section (Volume Two) contains commissioned papers on review systems that focus on schoolwide improvement models. The fourth section contains key documents related to the Expert Panel System, and the fifth section contains conceptual and planning papers in chronological order. The two volumes contain the following papers and citations from legislation and regulations: (1) "A Synthesis and Integration of U.S. Evaluation Efforts To Identify Promising and Exemplary Educational Programs, Products, and Practices" (Gary Borich); (2) "Examining Federal Approaches Outside the Department of Education To Identify and Disseminate the Best" (Karen Bogart); (3) "Notes on the Federal Dissemination Experts Meeting: 'Advising the Department of Education's Office of Educational Research and Improvement on Its Design of a Systematic Consumer-oriented Evaluation System To Designate Promising and Exemplary Educational Products, Programs, and Practices' Oct. 12, 1995, Academy for Educational Development, Washington, D.C." (Karen Bogart); (4) "Standards and Their Use in the Food and Drug Administration (FDA)" (Gerald Barkdoll); (5) "Review of Foundation, Associations, and Non-profits Practices in Designating Promising and Exemplary Programs" (Janet Carter and Diane Schilder); (6) "Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas" (John Luczak and Joan Ruskus); (7) "A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science" (Carol Muscara); (8) "Evaluation and Standards for Schoolwide Programs and Programs Conducted in Multiple Sites" (John H. Hollifield, Samuel C. Stringfield, and Rebecca Herman); (9) "Changing the National Diffusion Network To Address Whole-School Reforms"

(Susan Bodily and Thomas Glennan); (10) "Key Parts of the 1994 OERI Authorization Relating to Expert Panels"; (11) "Draft Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI)--Designation of Exemplary and Promising Programs, Proposed Rule, June 3, 1996"; (12) "Final Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI)--Designation of Exemplary and Promising Programs, Nov. 1997"; (13) "Developing and Operating Expert Panels: A Key Leadership and Coordination Role for OERI" (Susan Klein); (14) "Leadership in Developing a Nation-wide 'Findbest' Education R&D Evaluation System: A New OERI Responsibility" (Susan Klein); (15) "Technical Assistance and the Creation of Educational Knowledge" (Brenda J. Turnbull); (16) "OERI Leadership Efforts in Designing Approaches To Identify and Share Promising and Exemplary Products, Programs, and Practices: Presession to the 1996 Annual Convention of the American Educational Research Association, New York City, April 8, 1996) (meeting notes by Allen Schmeider); (17) "Responses to: Options in Discussing Distinctions between Promising and Exemplary Products, Programs, Practices, Policies and Research Findings for a System of Expert Panels" (Susan Klein); (18) "Expert Panels for Promising and Exemplary Innovations: A 'Fine Idea from the Feds'" (Peter Seidman); (19) "Design Competitions: A Proposal for a New Federal Role in Educational Research and Development" (Robert E. Slavin); (20) "Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches To Find, Develop, and Share Promising and Exemplary Products and Programs" (Susan S. Klein); (21) "The Evolving System of Expert Panels, a Presession on April 13, 1998 at the American Educational Research Association Conference, San Diego" (Chris Dwyer, Susan Klaiber, and Peggy Simon); and (22) "The Expert Panel System Is a Reasonable Way To Identify Promising and Exemplary Programs--a Response to Stanley Pogrow's Article, "What Is an Exemplary Program and Why Should Anyone Care? A Reaction to Slavin and Klein'" (Susan S. Klein). (SLD)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

ED 429 130

Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs

Working Papers

Volume One

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM029699

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

BEST COPY AVAILABLE

Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs

Working Papers

Volume One

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

Office of Reform Assistance and Dissemination
Peirce Hammond
Director

May 1998

Bound Edition March 1999

This publication is intended to promote the exchange of ideas among researchers and policymakers. The views expressed in it are part of ongoing research and analysis and, except for the official regulations, do not necessarily reflect the position of the U.S. Department of Education. Previously published documents such as articles from the *Educational Researcher* have been reproduced with permission of the authors and editors.

Note: In order to speed distribution of these working papers, we have reproduced original documents and their pagination.

Contact:
Susan Klein, Ed.D
(202) 219–2038
sue_klein@ed.gov

Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs

(Volumes One and Two)

Tab

VOLUME I.

Introductory Papers

Overview: Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share the Best -- Susan Klein, Office of Educational Research and Improvement (OERI) A

Preface: Michael Scriven, Claremont Graduate University B

Current Status of the System of Expert Panels: Sharon Bobbitt and Susan Klein, OERI (April 1998) C

Incorporating Research Findings and Practice in Expert Panel Work: A Dialog Between Michael Scriven and Lois-ellin Datta (February 1997) D

II. Papers Examining other Review Activities to Identify the Best

A Synthesis and Integration of U.S. Evaluation Efforts to Identify Promising and Exemplary Educational Programs, Products and Practices -- Gary Borich, Univ. of Texas, Austin (July 1997) E

Examining Federal Approaches Outside the Department of Education to Identify and Disseminate the Best -- Karen Bogart, Anne Steinmann Institute (June 1997) F

Notes on the Federal Dissemination Experts Meeting: "Advising the Department of Education's Office of Educational Research and Improvement on its Design of a Systematic Consumer-oriented Evaluation System to Designate Promising and Exemplary Educational Products, Programs and Practices" Oct. 12, 1995, Academy for Educational Development, Washington, DC. -- Karen Bogart, Ann Steinmann Institute G

Standards and their Use in the Food and Drug Administration (FDA) -- Gerald (Jake) Barkdoll (Jan. 1996) H

Review of Foundation, Associations, and Non-profits Practices in Designating Promising and Exemplary Programs-- Janet Carter and Diane Schilder (July 1997) I

Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas -- John Luczak & Joan Ruskus, SRI International (July 1997) J

A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science -- Carol Muscara, Computer Technology Services, Inc., Rockville, MD (Nov. 1996) K

Tab

VOLUME II

III. Papers on Schoolwide Models

Evaluation and Standards for Schoolwide Programs and Programs Conducted in Multiple Sites -- John H. Hollifield, Samuel C. Stringfield, and Rebecca Herman, Johns Hopkins Univ. (Jan. 1996) **L**

Changing the National Diffusion Network to Address Whole-School Reforms -- Susan Bodily and Thomas Glennan (Oct. 1996) **M**

IV. Legislation and Regulations

Key Parts of the 1994 OERI Authorization relating to Expert Panels **N**

Draft Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI) -- Designation of Exemplary and Promising Programs" Proposed Rule, June 3, 1996 **O**

Final Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI) -- Designation of Exemplary and Promising Programs, Nov. 1997 **P**

V. Conceptual and Planning Papers (in Chronological Order):

List of Papers and Publications Related to the System of Expert Panels **Q**

Developing and Operating Expert Panels: A Key Leadership and Coordination Role for OERI -- Susan Klein, OERI (Jan. 16, 1995) **R**

Leadership in Developing a Nation-wide "Findbest" Education R&D Evaluation System: A New OERI Responsibility -- Susan Klein, OERI (Jan. 29, 1996) **S**

Technical Assistance and the Creation of Educational Knowledge -- Brenda J. Turnbull, Policy Studies Associates (March 1996) **T**

OERI Leadership Efforts in Designing Approaches to Identify and Share Promising and Exemplary Products, Programs and Practices: Presession to the 1996 Annual Convention of the American Educational Research Association, New York City, April 8, 1996 Meeting notes by Allen Schmieder, OERI **U**

Responses to: Options in Discussing Distinctions Between Promising and Exemplary Products, Programs, Practices, Policies and Research Findings for a System of Expert Panels -- Susan Klein (Nov. 19, 1996) **V**

VOLUME I

Section I. Introductory Papers

Overview: Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share the Best -- Susan Klein, Office of Educational Research and Improvement (OERI)	**A**
Preface: Michael Scriven, Claremont Graduate University	**B**
Current Status of the System of Expert Panels: Sharon Bobbitt and Susan Klein, OERI (April 1998)	**C**
Incorporating Research Findings and Practice in Expert Panel Work: A Dialog Between Michael Scriven and Lois-ellin Datta (February 1997)	**D**

Tab

VOLUME II, Continued

Expert Panels for Promising and Exemplary Innovations: A "Fine Idea From the Feds" by Peter Seidman, NCRVE Centerwork: National Center for Research in Vocational Education, Univ. of California, Berkeley (Winter 1997) W

"Design Competitions: A Proposal for a New Federal Role in Educational Research and Development" by Robert E. Slavin, *Educational Researcher* (Jan./Feb. 1997) X

"Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop, and Share Promising and Exemplary Products and Programs" by Susan S. Klein, *Educational Researcher* (Aug./Sept. 1997) Y

The Evolving System of Expert Panels, A Presession on April 13, 1998 at the American Educational Research Association Conference, San Diego, Sponsored by the Office of Educational Research and Improvement, OERI, U.S. Department of Education. (Prepared by Chris Dwyer, Susan Klaiber and Peggy Simon, RMC) Z

The Expert Panel System is a Reasonable Way to Identify Promising and Exemplary Programs --- A Response to Stanley Pogrow's Article, "What is an Exemplary Program and Why Should Anyone Care? A Reaction to Slavin and Klein." (Susan S. Klein, October 26, 1998). AA

A

Overview: Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs

By

Susan S. Klein, Knowledge Applications Division, Office of Reform Assistance and Dissemination, Office of Educational Research and Improvement (OERI), U.S. Department of Education (ED)

This is one of several reports which we have developed to guide the new OERI facilitated System of Expert Panels. It contains OERI staff and externally developed papers which explore issues and activities related to the purposes of the expert panels -- to identify and facilitate the sharing of promising and exemplary products, programs and practices in specific topic areas where the Department of Education is expected to provide leadership. It is hoped that these expert panels will be long-term improvement efforts that are "owned" by the appropriate constituent groups in the topic areas and that the Panels will learn from each others experiences. Each Panel will have both short and long term goals. The short term goals will be to help identify the most viable R&D based strategies and models and describe their strengths and weaknesses so that educators will be able to make wise choices among them. In addition to making this consumer information available on a continuous basis, the longer term goals include:

- encouraging the constituents in a specific field to learn to invest in evaluations of replicable models so that they will have a continuous systematic way to learn what works and to use similar indicators so that the success of different models can be compared -- particularly as they are used with diverse populations and in different contexts.

- identifying promising models and strategies which need and merit additional investment in the improvement of their quality and utility and in adapting them for new types of users.

- locating gaps where the federal government or others may want to fund the development of new models perhaps using the design competition strategy suggested by Robert Slavin in "Design Competitions: A Proposal for a New Federal role in Educational Research and Development" 1997, *Educational Researcher* 26(1), 22-28 [Tab X] and discussed as it relates to expert panels by Susan Klein in "Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop, and Share Promising and Exemplary Products and Programs" 1997, *Educational Researcher* 26(6), 12-20 [Tab Y].

A long term systematic approach to this type of a consumer oriented review in specific topic areas is new for the Department of Education, although it builds on what has been learned from previous and current multi-topic federal education review efforts such as the Joint Dissemination Review Panel and the Program Effectiveness Panel. These education review panels are ably described in commissioned papers by John Evans and Christine Dwyer in *Disseminating Promising and Exemplary Programs: Planning a System of Expert Panels*, December 1997, the first in this series of OERI working paper volumes. This new System of Expert Panels also builds on some other short term single topic recognition and replication review activities described in, "Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas" by John Luczak and Joan Ruskus [Tab J].

Introductory Papers in Section I, are followed by commissioned papers on a variety of review activities designed to find the best in Section II. These review activities are conducted by federal agencies in and outside of education, foundations and associations. Section III contains commissioned papers on review systems which focus on school wide improvement models. Section IV contains key documents related to this System such as pages from the OERI legislation and the draft and final regulations for the standards on exemplary and promising programs. Section V contains conceptual and planning papers in chronological order.

In addition to these two reports on issues related to the System of Expert Panels, we are planning additional reports to document the pilot experiences of the Gender Equity Expert Panel and the Mathematics and Science Expert Panel.

Since these OERI volumes are published as working papers for a special audience of individuals who are interested in planning relating to this emerging system of Expert Panels, we have included them as we received them from the authors and not gone to the expense of retyping and giving them a consistent format.

B

PREFACE

Michael Scriven
Claremont Graduate University

This is intended to be a working preface, making a small contribution to the essays in this important volume. But perhaps I can lead into the constructive suggestions via a relevant fragment of historical background. The long road to this volume began many years ago when a contract was let by the U. S. Office of Education to the Educational Testing Service to select some of the products of the federal labs and centers for dissemination funding. So, one day in 1970 or thereabouts, a group of consultants, myself included, found ourselves in an office at ETS studying products and the supporting documentation. There was at that stage no precedent to follow, unless you reached as far as *Consumer Reports*.

As I began to identify recurrent weaknesses in the documentation spread across the floor, it became clear that we needed to create a checklist of items that should be covered. That checklist notably improved both our speed of operation and the reliability of our evaluation process. Some dozens of revisions later, the latest edition of that checklist still serves to organize most of the product - and program - evaluations I do today. Its strength lay in its emergence from the shortcomings of those multiple attempts by the applicants to make a good case for the merit of a particular product. I believe that provides a better - because it is a more reality-related - starting point than trying to think up a list of what makes a product or program good. In the history of logic, it is the fallacies approach, which has turned out to be much more useful than the 'mathematical ideal' approach. Certainly that approach provides an unpacking of some of the more complex notions in the present list of key criteria to be used by the new, multiple expert-panel approach to educational product quality control and dissemination - especially the notions of educational significance, exemplary, and promising.

The checklist I'm referring to - called for convenience the Key Evaluation Checklist (KEC or K-list) - has a total of 19 checkpoints, but the 'hard core' ones are numbers 5 to 11. They are:

K5. VALUES__(Needs, wants, professional standards, legal standards, ethical standards, scientific standards, etc.) Here is the primary location for principles of equity. This is the only source of the evaluative conclusions that are achieved by combining values with the factual results in K6 to K10.

K6. PROCESS__(Assessment of everything significant that is done before outcomes emerge.) Here one applies ethical considerations such as equity to the project's activities, e.g., to the selection procedures used to govern admission to entrepreneurship seminars. Another example: if the materials contain allegations about the superiority of some pedagogical approach to teaching scientific method (e.g., collaborative learning), the scientific basis for those claims must be solid.

K7. OUTCOMES_(Impact on consumers: direct/indirect, intended/unintended, immediate/short-term/long-term.) Here one applies the needs assessment as well as the equity principles of K5 to the results, intended and unintended, from the project. Note that outcomes are not the same as outputs; for a project that produces educational pamphlets, those are the output, but only the effects on readers of the pamphlets count as outcomes (a.k.a. impact). Outputs, however high their quality, have no educational significance: only outcomes count, in the hard currency that can be used to improve the world we live in - and that is the currency of promising and exemplary products and practices, since to qualify as these, good results (outcomes) must have been demonstrated.

K8. COSTS_(Money, non-money; direct, indirect, and opportunity; start-up/maintenance/upgrade/shutdown; time-discounted and inflation-corrected; etc.) A key consideration in determining educational significance; a materials project for the social studies classroom whose product will cost a school more than 90% of the materials for K-12 classrooms (i.e., more than, roughly, two dollars) is unlikely to have much educational significance - of course, this rules out essentially all educational software unless it is extraordinarily successful.

K9. COMPARISONS_(Other ways of getting the same or more benefits from available resources.) Innovations that prove no better, on overall merit or cost effectiveness dimensions, than existing products/practices (or obvious combinations of existing practices), have no educational significance.

K10. GENERALIZABILITY/REPLICABILITY/EXPORTABILITY (The dimensions of generalizability include not only useability at other sites, but by other personnel, at later dates, with other content, and in other types of program; replicability and exportability usually refer to other sites.) The exemplary product has already scored on this checkpoint, along one key dimension of export; the promising product has shown it is a legitimate entry in this race. Consider rating it still higher if it seems likely to succeed in other dimensions of generalizability. The product whose only significance lies in its interest from the point of view of theory has not yet shown itself to be promising, as that term is used here. Few entities of theoretical interest make the transition to reality-value, and since even experts are not reliable in judging which ones will do so, it is inappropriate to spend money on them until they pass the initiation test of single-site payoffs. Keep in mind that products with quite a modest impact at one site may be highly significant if they can be successfully introduced at a thousand sites.

K11. OVERALL SIGNIFICANCE __(The sum of the above, focused on the consumers' needs.) From the KEC viewpoint, significance is not something that can be directly judged: it is instead a complex compound of K5 to K10, a matter of looking carefully at each of the performance features (K6 to K10) and then at whether this packet of properties matches the needs from K5. Often, the cost level or comparisons checkpoints alone will completely knock out an otherwise very attractive entry.

Now the KEC does not do the whole job of evaluation. In the first place, when evaluating certain types of products such as computer software, one needs some area expertise to develop more detailed checklists in order to come to a conclusion about the merits of a product or its effects. (The KEC is a macro checklist; one also sometimes needs micro-checklists.) Then of course there is the matter of determining the true situation with respect to the performance of an entry on all the dimensions of merit, from the macro to the micro, and at one or several sites. And finally there is the tricky business of making a solid inference from the constituent performances on K6 to K10, to the synthesis represented by K11. If one wants a ranking of entries, one needs some way to represent an overall score on significance, for example, some rubric for generating an overall grade for the product from the grades for the checkpoints. Of course, one can just eyeball this, which is what panels generally do, but that is a practice that leads to inconsistency.

Finally a word about the future. We have made a fine beginning with the panels on gender equity and science/mathematics, which relate to truly important elements in education. Along the road ahead, we will need to apply the panel approach to products in such areas as reading and numeracy and computer competencies. My sense of the importance of the task can be suggested by the following rough estimates in one of these cases. The thin evidence we presently have suggests that the best available materials for teaching beginning reading are very much more effective than any of the four or five most widely used materials. This edge in effectiveness translates into the ability to convert failure into success for some large proportion of the students that do not now become functionally literate. Suppose we are now seeing 30,000 students leave the schools each year without becoming functionally literate, an estimate based on the NAEP and other data. If this can be cut by a third as a result of switching to the best products available, as I think is readily possible, that means we could save 100,000 students from illiteracy every decade. Given the savage restrictions that illiteracy imposes on vocational options, we can see that serious educational product evaluation quickly makes possible major gains in quality of life and voting effectiveness for many of our future fellow-citizens.

For more than half a century this country has been a leader in providing useful product evaluation to purchasers of drugs and dentifrices, cars and can openers. When the stakes increase to matters of salvation and success for our children, there is surely no excuse for further postponing the delivery of high quality educational product evaluations.

February 18, 1998

C

Current Status of the System of Expert Panels

By Sharon Bobbitt and Susan Klein
Knowledge Application Division, Office of Reform Assistance and Dissemination, Office of Educational Research and Improvement, U.S. Department of Education

Background

On March 31, 1994, President Clinton signed Public Law 103-227, which includes Title IX, the "Educational Research, Development, Dissemination, and Improvement Act of 1994" (Act). The Act restructured the Office of Educational Research and Improvement (OERI) and provided it with a broad mandate to conduct an array of research, development, dissemination, and improvement activities aimed at strengthening the education of all students. The legislation directed the Assistant Secretary, in consultation with the National Educational Research Policy and Priorities Board (Board), to develop standards to govern the conduct and evaluation of all research, development, and dissemination activities carried out by OERI to ensure that these activities meet the highest standards of professional excellence. Known as the Phase II standards, they were published in the *Federal Register* on November 17, 1997. (See <http://www.ed.gov/legislation/FedRegister/finrule/1997-4111797b.html>).

The OERI legislation also requires that expert panels be established to review educational programs and to recommend to the Secretary those programs that should be designated as exemplary or promising. The legislation defines "educational program" to include "educational policies, research findings, practices and products." These programs should be disseminated through the Department's National Education Dissemination System (NEDS), the Educational Resources Information Center (ERIC) Clearinghouses, the regional educational laboratories, and other like entities. The legislation further requires that the Assistant Secretary develop standards that describe the procedures the panels will use in reviewing educational programs. For an educational program to be considered for designation as exemplary or promising, the eligible entity must submit to the Secretary a description of the program, program materials, and a discussion of the program that is responsive to the established criteria.

After a review of the public comments on the Phase II standards, it was agreed that for the next year or two it would be best to keep the standards as general as possible. Since the expert panels and program reviewers are composed of experts in the field, the general thinking was that the panels themselves would be competent to establish criteria according to their own disciplines.

OERI established two pilot expert panels, one on mathematics and science, the other on gender equity. These two panels have established procedures for receiving, evaluating, and--to some extent--disseminating exemplary and promising practices in their respective domains. Submission guidelines for these panels are available on their respective Web sites ---<http://www.rmcres.com> or <http://www/edc.org/WomensEquity/pubs/expanel/paneldescr.html>and<http://www.edc.org/hec/genviol/>

Mathematics/Science Expert Panel

The math/science panel is composed of 15 experts in mathematics or science from around the country. It includes academics, association representatives, regional lab representatives, and practitioners. The panel met five times during the past eighteen months to establish procedures for submissions, criteria, and rubrics that the reviewers can use to evaluate the submissions. The criteria flesh out the standards categories outlined in the Phase II standards, namely, quality of the program, educational significance, replicability, and evidence of effectiveness or success.

Programs that are recommended as promising do not have to meet the rigorous standards for exemplary programs. Part of the purpose of the legislation was to increase the number of good educational practices available to the field. The math/science panel recognizes the need to make a wide variety of good programs available to states, districts, and schools. Programs designated as promising have a role to play in meeting this need. They will, however, be designated as promising for only a five-year period and then either meet the criteria for exemplary programs or be removed from the list of promising programs. (Under extraordinary circumstances, a program may continue to be labeled as promising for up to 3 additional years.)

During the summer, the panel conducted two pilot program review tests, one in science and one in mathematics. These pilots had several purposes, including to test the numbers of reviewers and types of expertise needed, and to examine the usefulness of the rubrics. Based on the results of the summer pilots, the math/science panel put the review process and the criteria into final form and is now in the process of making decisions on over 60 mathematics submissions received by the Jan. 30, 1998 deadline.

The Federal Coordinator of this Panel is Patricia O'Connell Ross who manages the Eisenhower Professional Development Program in OERI's Office of Reform Assistance and Dissemination (ORAD). She may be reached at 202-219-2169 or <patricia_ross@ed.gov>.

The Gender Equity Expert Panel

The pilot of the Gender Equity Expert Panel started its planning work in the fall of 1995. The Panel held an initial, partial panel meeting in April 1996 and another meeting to work on recommendations about 20 submissions in Sept. 1997. The support contractor for this panel has been the Women's Educational Equity Act Equity Resource Center at Education Development Center (EDC) which also includes information on the Panel in its web site. The Safe and Drug Free Schools Program is funding the subpanel on the Prevention of Sexual and Racial Harassment and Violence against Students in Higher Education out of its higher education center administered by EDC. The deadline for these submissions in May 1, 1998. The Office of Special Education and Rehabilitation Services (OSERS) is funding the subpanel on Gender Equity and Disability. The four other subpanels: Mathematics, Science and Technology; Teacher Education and Professional Development; Vocational Technical Education/School to Work; and Core Gender Equity have not had funding for subpanel meetings.

OERI staff, the support contractor, and the subpanel chairs, with help from other panel members and the 100 plus member group of interested colleagues, have developed submission and review guidelines based on the draft OERI Standards for Exemplary and Promising Programs. As part of its pilot phase, the 30-member panel actively solicited nominations from the field and obtained 20 full submissions by the due date of July 1997. The support contractor and the subpanel chairs selected reviewers for each submission, each of whom had relevant expertise and no conflicting interests. The professional background of the reviewers ranged from equity specialists to researchers to teachers. Some of these reviewers were also Panel members. The reviews were synthesized in decision memos by each subpanel for discussion at the Sept. 12, 1997 meeting of the full panel. The full panel agreed that more information was needed to make final decisions. As a consequence, the subpanels are in the process of working with the submitters to obtain additional information and with OERI staff and members of the Impact Review Panel to further clarify the Panel decision rules. To comply with the Phase II standards and also to meet the needs of their constituents, the Gender Equity Expert Panel has decided to make three types of recommendations:

1) recommend follow-up dissemination, revision or evaluation activities for **information products** such as research syntheses, videos, and college course materials that increase user awareness and knowledge of gender equity in ways that reflect the criteria categories of quality, educational significance and usefulness to others;
2) recommend **interventions** as *promising* if they are adequately responsive to the above three criteria categories and also have some evidence that the intervention has increased gender equity in at least one context; and
3) recommend **interventions** as *exemplary* if they are positively responsive to all four criteria categories including evidence of success or effectiveness. Claims of success in promoting gender equity must demonstrate that the positive results can logically be attributed to the intervention and that they are fairly consistent in appropriate multiple settings.

The panel has presented at annual meetings of AERA, the National Coalition For Sex Equity In Education, and written extensively about the Gender Equity Expert Panel to guide its development and to help interested colleagues participate in Panel related activities. Susan Klein, in the Knowledge Applications Division of ORAD is the federal Coordinator and conceptualizer of this Panel. She may be reached at 202-219-2038 or <sue_klein@ed.gov>.

Safe and Drug-Free Panel.

The Office of Elementary and Secondary Education is funding a panel on Safe and Drug-Free Schools. The Safe and Drug Free Schools program seeks to identify and disseminate promising and exemplary programs, especially among its grantees. The panel is slated to review up to 80 submissions during this fiscal year. Ann Weinheimer, from the program office, serves as the federal Coordinator for this Panel. She may be reached at 202-708-5939 or <ann_weinheimer@ed.gov>.

Educational Technology Panel

Plans for the technology panel call for the establishment of the panel this year. The Office of Educational Technology has requested that we accelerate the establishment of this Panel. We expect to have the first panel meeting in late summer of 1998. Cheryl Garnette and Diane Aleem in the Learning Technologies Division, ORAD are the federal Coordinators of this Panel. Cheryl may be reached at 202-219-2267 or <cheryl_garnette@ed.gov> and Diane at 202-219-2148 or <diane_aleem@ed.gov>.

Reading Panel

Current plans call for the establishment of an expert panel to look at reading this year, in line with the President's initiatives. To help support the initiatives in early reading and pre-school education the initial focus of this panel will be on reading preparation and skills for young children. OERI plans to establish the panel in the summer of 1998. Naomi Karp, the Director of OERI's National Institute on Early Childhood Development and Education, is the federal Coordinator of this Panel. She may be reached at 202-219-1586 or <naomi_karp@ed.gov>.

The Impact Review Panel

The math/science panel recommended that a panel of evaluation experts review their submissions to see if the results provided are valid before a program receives an exemplary designation. This idea has been adapted and applied to all Panels to ensure consistency and reliability in the decisions on the programs that the Panels recommend that the Secretary of Education designate as exemplary. As a result, all the topic focused expert panels will use this Impact Review Panel (IRP) of evaluation experts to obtain advice on whether each individual program being considered as exemplary has evaluation design and evidence to support meaningful claims of success. The IRP will examine whether the claims are supported by evidence that is logically and, as appropriate, statistically valid. The first meeting of the ten member IRP was April 13, 1998.

Personnel

Peirce Hammond, the Director of ORAD, and Sharon Bobbitt, the Director of ORAD's Knowledge Applications Division (KAD) provide overall support and leadership for this new System. Within KAD, the Acting Team Leader for the System of Expert Panels is Susan Klein. Stephen O'Brien serves as COTR for the System contract with RMC Research Corporation of Portsmouth, NH. Susan Klaiber serves as the project director and Peggy Simon as assistant director. They can be reached at (800) 258-0802 or <sklaiber@rmcres.com> or <psimon@rmcres.com>.

April 28, 1998

D

20

Incorporating Research Findings and Practices In Expert Panel Work:
A Dialog Between Michael Scriven (MS) and Lois-ellin Datta (LD)

THE BACKGROUND

Legislation reauthorizing the U.S. Department of Education's Office of Educational Research and Improvement (OERI) cuts a wide swath in what may be designated as promising or exemplary. The expert panels may recognize not only products and programs, but also educational policies, practices and research findings. Precedent exists for harvesting each of these elements.

As examples.

o for almost three decades, the ERIC system has been screening research reports, gently for inclusion in its archives and more tough-mindedly for inclusion in its state-of-knowledge publications.

o the National Institutes of Health's (NIH) Office of Medical Applications Research (OMAR) panels meet about four to six times yearly to determine what works best for treating problems such as prostate cancer, back pain and ulcers.

o OERI-funded labs and centers are authorized to prepare state-of-the-art reviews of research in their areas, which requires scrutinizing research quality

o third-party groups have received federal funding to develop systems for and conduct reviews of educational products

o the Blue Ribbon Schools and other recognition efforts such as Teacher of the Year are nomination-based "promising and effective" systems for educational organizations, persons and through them, practices

These efforts have co-existed peacefully for many years. There is no direct mandate to bring them under the same umbrella of concepts, standards or procedures for recognizing quality. However, *indirectly*, the OERI mandate for the panels could set an intriguing precedent. This may be the first time that a systematic effort will be made to talk about and possibly develop common guidelines for recognizing excellence in all five "educational solutions" : research findings, policies, practices, products and programs.

OERI has requested (October 2,1996) a discussion of some aspects of incorporating research findings and practices in panel work. The request takes the form of four questions (reorganized and paraphrased from the original request) : (1) should panels support development of research syntheses, (2) should panels review research itself for designation as promising or exemplary, (3A) should panels explicitly attend to existing standards or laws which were based in part on research findings, (3B) should national or state mathematics and science education standards be used to determine program quality; and (4) should panels treat the five types of educational solutions in the same way?

In the next section, each question is considered. The paraphrase is followed by the original OERI phrasing.

In these comments, "research" is taken to mean "..diligent and systematic inquiry into a subject in order to determine or revise information about facts, theories, applications, etc." (Webster International Dictionary). Evaluation is included as research. It is a form of systematic and diligent inquiry, one specially concerned with establishing merit, worth and value, using Scriven's definition, "..the process of determining the merit, worth or value of something, or the product of that process" (Evaluation Thesaurus, 1991, p. 139)

THE QUESTIONS

1. The role of research in panel activities: should panels support development of research syntheses?

"When, if ever, should an expert panel support the development of research syntheses or a review of these syntheses in a given topic area to identify those that are the best for subsequent use of expert panel reviews and others?....Should panel procedures be guided by the philosophy that if the panel members themselves have the appropriate expertise, they will know good claims and good evidence when they see it?....Should research findings and syntheses be viewed by expert panels more as guidance for criteria relating to good indicators of success or aspects of quality criteria, than as the 'disseminands. ' If yes, the system of expert panels could explain that they are not going to be reviewing individual research findings or reports to determine if they are promising or exemplary, but that what is learned about these collective findings will be used 'instrumentally' in decisions about products or programs." (OERI, Oct. 2, 1996)

LD Yes , the expert panels should use research syntheses for information. No, they should not rely wholly on their own expertise. Yes, the research syntheses they are using should be a matter of public record and available to applicants, evaluators and researchers. Yes, if there is good information about indicators of success, this should be made available to panelists and others.

That is, the expert panels should be able to commission research syntheses as needed and should be guided by the philosophy that even experts have something new to learn.

An Anecdote: In the Joint Dissemination Review Panels, year after year, eager applicants would breathlessly report the effectiveness of a brand-new approach: individualized instruction. The applicants often showed no awareness that others had been there before and few looked right or left to see how others had individualized instruction or whether their own approach offered any value-added to existing approaches. It was enough for these applicants to state that their project goal was increasing reading or mathematics achievement (or self-esteem or school attendance) and by gum, relative to the baselines in September, the children were better in May...or showed significant increases over gains during the previous year or scored higher than the lowest possible norm group.

These projects came to the Panel, it should be remembered, only after experts at federal, state and local levels had selected the applicants as showing exemplary effectiveness. And the applications often had been written by those developing and evaluating the programs.

This anecdote can be written off as long ago, far away and in benighted times. In our times, however, the panels which are the gold-standard for carefully assessing evidence of treatment effectiveness (the National Institutes of Health OMAR panels) *begin every review* with a first-class, extensive, independent review of the literature—a research synthesis. This is not taken as casting aspersions on the expertness of the experts on the OMAR panels. They are assumed to be exceptionally well-informed, the brightest of the bright, the most authoritative of the authorities. It is assumed, however, that almost no one can be fully up-to-date on a body of evidence and experience, and assumed further that calibration of panel information is a starting point for their deliberations.

Panel members may know good evidence when they see it. They may or may not know —in a systematic way—how the evidence they are seeing and the claims they are hearing stack up against the state of the art.

In any major area where the panel is reviewing a submission, all panel members should have the benefit of access to a state-of-the-art summary of relevant experience and research in the area. If a first-rate recent research synthesis already exists, it could be distributed to panel members. If one needs to be commissioned, the Panel (or OERI, or the panel support contractor) should have the power, money and time to have such a synthesis prepared.

There is no way to force panel members who think they know it all to read a synthesis—and given the overwhelming demands for free labor so well documented in Pat Campbell's formative evaluation, this might be the final straw in panel participation—but at least OERI will have done its best to assure a submission is read in the context of what else is happening, what else is known.

There is one non-trivial fine-line here. The panels should be a venue for iconoclastic, original, outrageous thought on how to do it better, approaches that break the mold, that may go counter to established beliefs or challenge political correctness. The research syntheses should not become a template for screening out programs whose assumptions and approaches do not conform to current beliefs.

The focus must be on results---not process. That is, the effectiveness criterion is not whether the approach uses the processes the panel thinks should work. The criterion should be results: sound evidence that the approach DOES work.

Extrapolating from previous research and preconceptions about what oughtta work, should work, might work is no substitute for evidence that something actually works. "Congruence with sound research and previous practice" should not be among the quality criteria in a system seriously seeking effective new ideas.

MJS This position, well expressed and supported by LED, might even be strengthened. To do so, it is first necessary to clarify the concepts involved. The panels are not only identifying meritorious practices - their explicit mission - but also, reading the context, they are (indirectly at least) making recommendations for the distribution of scarce resources. These resources include: (I) dissemination funding, and (ii) the sustained time and attention of teachers and teacher trainers. It is in this context that one needs to interpret "exemplary" (E) and "promising" (P). Hence it would be inappropriate for them to identify as E or P (E/P) a project that is already well known to be successful and is widely used, e.g., the use of wait time, or computer simulation of dangerous chemical experiments or of anthropological observations in the rain forest.

It would be redundant to support the dissemination of information about such a practice. One possible exception to this, which is considered below, might arise in appraising a practice which is now being ignored although the evidence is strong that it was and would - if used - continue to be very successful. Barak Rosenshine has collected a substantial list of these - an example would be the use of programmed texts in an appropriate context. In general, however, the mission of the panels, understood in their working context, is to identify and support new (or refined or extended) E/P practices. It follow that their mission would be hard to discharge effectively if based on only a vague idea - or on multiple conflicting ideas - of the state of the art. This account would suggest that the expert panels might well begin by pre-reading existing meta-analyses, and then discussing them - and their own overviews - at their first meeting. Where prior meta-analyses are not available or not judged satisfactory (and if they lack agreement on their own meta-analysis), they might decide whether to commission a new one. (In some circumstances, thought might be given to commissioning two of these, to be done independently, i.e., without consultation between the analysts.)

At first, where it is necessary to proceed faster than is possible by awaiting the completion of new meta-analyses, the panel could of course go with their own impressions, even if inconsistent; in the latter case covering their bets by backing some very different approaches. But the key point would be to realize that meta-analyses would contribute usefully to the panels' work, and commission them so that they would be used at later meetings or by later panels. We cannot accept less in education than the NIH paradigm, which, as LED rightly reminds us, does exactly this - always commissioning meta-analyses at the start of a new major review. The costs of not doing it are to be seen in the historical data on the great power of fashion swings in education, often thinly disguised by a name change, even when the evidence has not changed.

At this stage it should be stressed that the issue is not whether there are meta-analyses of _any_ kind in existence, but whether they are of a particular kind, namely _evaluative_ meta-analyses of interventions ("practices"). Many meta-analyses are simply reviews of research, much of which is theory-oriented; what are needed by evaluative panels are reviews of evaluative research on practice. (Example of the difference: research may discover strong relationships between SES and academic performance, but this may have as little significance for educational practice as the personal advice "Be born of rich parents".)

What about the case of a project submitted to an expert panel that revives some now unfashionable practice, for which there was always quite good evidence of success? It is very important that the expert panels not endorse the swings of fashion as if they were swings of evidence. The panels work in a context of what is seen by practitioners and change agents to be new: they are not historians of science who would be compelled to reject the practice from the domain of originality. So, working in this context, the panels would be entirely entitled to identify as E/P something which is no longer accepted as part of the body of worthy practice, if its merits meet the requisite standards for current E/P practice.

2. Should the expert panels review research itself to determine if the research is promising or exemplary?

"Is it OK for expert panels not to review purely informational resources such as research reports, research syntheses and research interpretations or other conceptual articles for designation as promising or exemplary? The rationale is that such resources are unlikely to have evidence of effectiveness and thus would not be candidates for designation as promising or exemplary. If this decision against reviewing purely information materials is changed, would the criteria/standards need to be changed?" (Source, OERI, October 2, 1996)

LD: The panels should not review research itself to determine whether the research findings are promising or exemplary. It is OK to pass. To the reason given by OERI for the panels' not reviewing research (such resources are unlikely to have evidence of effectiveness), two other reasons are added: avoiding duplication and practicalities.

With regard to duplication, there are appropriate venues for recognizing promising and exemplary research and exemplary researchers : acceptance in top refereed journals, invitations to invisible college conferences, reviews of final reports and applications for new research grants, dissertation committees, the Nobel Prize, election to Fellowship status in professional societies, and the opinion of scientific peers, to mention a few.

More could be imagined. But adding this responsibility of additional recognition to that of the infant expert panels is unwise.

With regard to practicalities, if the panel members feel underpaid and overworked just talking about the panels before they have reviewed a single submission, attention probably should focus on educational products, programs, practices and policies.

One exception is reviewing claims about the effectiveness of new ways of going from knowledge to educational improvement. For example, one group may claim that a newly developed approach to evaluation utilization (dissemination, knowledge utilization) works better than the average bear. Another group may claim to have completed the best research ever on knowledge utilization, as pathfinding as the Iowa Seed Corn study.

The first claim ("We can demonstrate that using this approach, knowledge goes swiftly from thought to action.") is appropriate for panel review, using the same criteria as any other claim but with appropriate knowledge of what is convincing evidence in the field of knowledge utilization.

The second claim ("This is first-rate basic research.") should not be reviewed by the panel for reasons already given of duplication and practicality, plus the considerably greater scientific expertise required to make such a judgment and the time often required to distinguish the signal of a great study from the more ephemeral noise of not-so-great studies.

MJS: Here I read the enabling charge to the panels rather differently. On my reading, it is the case that the panels and the whole system are in fact specifically charged to review research. However, the practical impact of this is not overwhelming, because –in the context where the panels will work– the justification for promoting research results by federal dissemination

support will be rare. I follow LED's first reason - avoidance of duplication- quite closely in what follows, but introduce another exception which suggests one should not at this stage alter the policy:
The system of research journals is well established, and provides quite well for the dissemination of research results including some meta analyses, so there is less need for mere informational dissemination in the case of research results than in the case of practices. However, circumstances may arise when an exception would need to be made, e.g., because of the delay in the publication of research results in certain journals, some results (including negative results) might deserve fast dissemination to the field.

LED provides another, persuasive, example of an exception to the exclusion of research; hence it seems unwise as well as unnecessary to change the terms of reference. This consideration seems more weighty than the others advanced, to which the following brief comments are addressed.

(i) The OERI suggestion of October 2, that "..such resources are unlikely to have evidence of effectiveness and thus would not be candidates for designation as promising or exemplary.." seems shaky because there are often publications, especially in evaluation journals, which review practice. (ii)

The second comment of October 2 raises the concern that the standards would have to be changed to allow consideration of research results. I think not, except possibly in minor verbal respects. (I am fairly confident of this, partly because I have always been acutely aware - and supportive - of this dimension of the charge to the panels and would have protested any earlier exclusion of it.)

(iii) LED's second reason concerns the practicalities of overwork. Here it seems possible that staff at OERI could do most of the filtering out of research candidates, using the simple criterion mentioned above: is there any reason why this result should not be disseminated via the usual information-disseminating mechanisms? A short (but extendible) list could easily be provided of exception-generating criteria: urgency, common ignorance or misapprehension because prior publication was only in some arcane journal, etc. Cases raising new questions would then be referred to the panel for a decision as to whether a new criterion should be added to the list (or whether one of those already on it should be removed).

3A and 3B: Should the panels attend to existing standards or laws which were based on research findings and political consensus? Should the national or state mathematics and science education standards be explicitly used to determine the quality of the program to be reviewed?

LD: It depends. More specifically, it depends (a) obsolescence and (b) the proportion of research to the proportion of political consensus.

Where national or state standards are based on first-rate state-of-the-art research, they may be sensible proxies for other information about quality. It would be silly for panels to ignore this information.

Where national or state standards are obsolete because they are based on research replaced by newer, better research or development, they are not sensible proxies for quality. The more recent information should be used, not obsolete standards.

Where national or state standards are mostly political consensus with a dollop of research, they should not be used as a proxy for judgments about quality. As noted above, one concern is that the panels, particularly in charged areas such as gender equity and English, retain capacity to recognize iconoclastic approaches that may work better than what is politically acceptable practice. My impression is that the math and science standards were based on considerably more than a dollop of research while efforts to formulate standards in English shipwrecked on political shoals.

National standards should be used by the Panels where the national standards are based primarily on research and where this research has not be overtaken by better knowledge. Some states may have better standards than the national standards. It is a major chore even for the stout-hearted, however, to figure out whether Hawai'i's statement on one point means the same thing or something different from Oregon's for standard after standard after standard.

Wherever possible, set the cross-bar low for quality as defined by standards based on political consensus. Try to give an approach that may not be what a panel member would do a chance to be reviewed for effectiveness, as long as it is not straight out illegal (even with waivers) or obviously harmful.

Provide consumer information in a factual way about where a program does or does not conform to national standards.

MJS: Cutting to the chase, this means that they should not be given any weight as such but only when they reflect good evidence. It may be helpful to pick up some examples in more detail, examples that show one could not adopt a general principle of support. There is, for example, a marked difference between the quality of the NCTM standards and the NRC/NAS standards. Even with our longer perspective on NCTM, they still look sound enough, at least in principle; but the macro standards in the NRC document contain blunders which will bedevil us for years to come, e.g., with respect to the omission of computer science, and the disrespect for technology - especially the failure to recognize the non-dependence of technology on science (these and other problems were detailed in my review of the draft for OERI, only some being corrected in the final version). It is important that the U.K. Standards avoided some of these problems, especially with respect to technology. And even NCTM is not immune to problems; for example, it is becoming more obvious that the NCTM standards did not anticipate the extent of the problems about the extra requirements on teacher time required by new approaches to pedagogy and assessment. They probably could not reasonably have anticipated the size of these problems, another reason why one should not treat existing standards as enduring criteria of merit. The NAEP standards in mathematics ran into some other problems, serious problems that led to major misinformation of the public about the success of their children, mainly because of faulty direction by the governing board. So it is clear we should not shackle the panels to any pre-existing standards: they should use their judgment as to which are worth considering.

LED makes three other points which I'd like to support. (i) The problem of interpreting these standards. It would not be a worthwhile use of panel time to become involved in exegetical minutiae, so they should not have a primary role. (ii) LED warns us about the intrusion of the politically correct into these matters: the panels should try to bypass such matters as much as possible and indeed encourage new and non-paradigm approaches. (iii) LED concludes by saying that where possible some comment might be made about consistency or inconsistency with existing standards. I would only add that this should be done with caution so as to avoid the time on hermeneutics that might be required in order to be sure about such views.

4. Should expert panels treat the educational solutions in the same way?

"Should expert panels have separate or integrated reviews, and separate or integrated consumer-oriented reports, for these five types of educational solutions? What is the difference between an expert panel review of individual submissions and their review of a fairly comprehensive set of submissions with similar purposes? Separate consumer reports that just focus on teacher practices, institutional policies, or on products and programs have been developed in the past, but is this the best plan for the expert panels and subpanels? How feasible and desirable are integrated reviews of multiple products, programs, policies and practices addressing the same goals together to facilitate comparative judgments and to prepare for the development of a consumer report on the 'solutions' available for each topic? Is it possible that different emphases may make sense for different panels or even subpanels? When should the reports on the promising and exemplary resources be separate or combined? How should the review strategies foster participation from the field of producers and users? To what extent does it make sense for consumer for the reports of the review decisions to be combined in some ways (such as the reports from AHCPR) or separate for specific purposes such as for research syntheses or the development or research and practice based consensus standards and benchmarks? If one of the products of expert panels is to be consumer reports, should we have some general guidelines on these reports and how they are developed? When should there be a critical mass of submission that merit comparative review? How much agreement is needed for decisions? How important are face to face meetings of reviewers versus other ways of obtaining agreements and discussions among reviews? Is it important to have some consensus in approving the consumer reports as well as in the individual product or program reviews." (Source of all the above questions: OERI, October 2, 1996)

LD: Were I the Goddess of The Expert Panel System, each year one or two burning questions asked by teachers, parents, principals would be identified. A nation-wide call would go forth for one page submissions of exemplary and promising anything—policies, practices, programs, products---with evidence showing effectiveness with regard to educational solutions for these issues. Every applicant whose submission passed the one page screen would receive $1,000 to pay for development of a fuller submission. The panels would look at submissions by type of educational solution (policies, practices, programs, products and all-of-the-above approaches) using comparable criteria, particularly with regard to outcomes, results, effectiveness. Their task would be to judge merit and to rank what was best (not necessarily 1 to 100 but what made sense, such as fantastic and not fantastic). Assuming good evidence of an effective educational solution was found, the winnahs would be interviewed on Larry King Live, Ophra and Jim Lehrer.

Lacking this, I recommend the panels try the closest feasible approximation.

a. One difference between panel review of single submissions and review of a fairly comprehensive set of submissions with similar purposes is the panel would be able to judge merit and assign rank. The claim could be that something is worthy and also the best of the population of approaches to a common problem. This is likely to be more useful for the teacher, parent, principal, school board member, etc.

b. Focused or consolidated consumer reports serve different purposes. With computer archives, an either/or choice may not be necessary. Some consumers will want best teacher practices, or best policies, or best products, etc. Some will want best solutions (including combining all of the above) to a problem. Have it both ways.

c. Some other groups prepare solution focused reviews. Many other groups do integrated reviews of multiple products. I do not know of any systematic study of the relative effort, costs and benefits of individual submission versus integrated reviews. Assuming the same number of educational solutions were to be reviewed, if money were available to bring panels together, it probably would be as efficient to conduct integrated as separate reviews. If OERI is relying on free labor, as journal editors do for their reviews, one submission at a time may be all an individual panelist can contribute.

Maybe OERI could explore how other groups such as the one in Santa Fe involving nationwide panels get lots of donated time to review video products for children. For example, one approach may be to recruit graduate seminars so panel work gets better integrated with training and more reviewer-power is available.

d. Whether some panels or subpanels do integrated versus individual reviews for different types of educational submissions may depend more on capacity to do population searches and panelist willingness to donate considerable time than anything intrinsic to the different types of educational solutions.

e. Reports on promising and exemplary resources should be separate. As Michael has wisely noted, when it comes to grant money, the best have a better right to scarce resources. This is true not only in competition for federal money, but also in the opportunity costs for local adoptions.

f(1): The process should start in the grassroots with teacher, parent, principals specification of what THEY want to know and cycle back, giving them the answer in THEIR terms. Perhaps one reason why doom and gloom can surface about educational change and knowledge utilization, is that teachers may ask questions such as "What is the best way to teach spelling?" and they get back research syntheses on language acquisition. A school board member asks, "What's the best way to measure reading competency in terms of grade level equivalents" and gets back reasons why you shouldn't measure reading competency at all, but should assess elements or skills that can be used to improve individual instruction. Make information available in the various ways people ask questions (see for example, the on-line medical and health information approaches).

f(2): The review strategies should foster a sense of participation from the field of producers and users. To my mind, one way to do this is to find out what questions they want answered, make it easy for people to prepare the initial submission; send out the call for nominations widely; give incentives to nominate; and PROMPTLY get information back to participants on what happened—question lists, nomination information, results.

g. Yes there should be general guidelines on producing consumer reports and how they are developed.

h. If the comparative review approach is selected, one can not wait for submissions to make it happen. First, submissions should be reviewed quickly. Second, the comparative approach requires some effort to reach the population of eligible approaches to similar goals. Hopefully, N will be more than 1.

I. Judging by the brouhahas from AHCPR, close to 100% consensus is required, including publication of dissenting comments.

j. RAND probably has the best data comparing face-to-face panels with other ways of obtaining agreements and discussion among reviewers. Brooks' studies should be completed by now.

k. One of the strengths of the NIH consensus process are the safeguards keeping the panels insulated from the political process. The reports are a panel product, not an NIH product. NIH makes sure the panel actually writes the report, everyone signs it and there is an immediate press conference held by the panel to release the panel findings. However, the various forms of public communication in the multiple journals, etc., as far as I know are the handiwork of communication folks with review by panel members. As noted in "A Matter of Consensus," it isn't clear if ED can keep the process apolitical, but the NIH approach seems worth emulating.

MS: Ah, the deluge! At last the details of practice catch up with us! For me, the only way to handle all this is from a clear vision of what the great needs are that we are trying to meet. The problem here is not a particularly formidable one: when I took on the evaluation of all U.S. Office of Education projects in order to select those that would be given to the newborn NIE, we faced all these problems in a more acute form (because of the huge investments already made - many millions of dollars in some projects - in what turned out to be indefensible projects on their own merits), but once the mission was clarified, the problems came out in the wash. In this case it does seem easy enough to formulate the mission: "Identify the best ways to help students learn the most valuable things we can help them to learn" - and the rest of the problems can be handled simply enough, by saying "Pick your best shot for the first round, without agonizing about it - the only important feature is that you now set up the machinery for serious review and improvement after the first round... And again after the second round... And again when it gives trouble or someone comes up with a better suggestion."

In what follows, I answer each of the questions raised about procedure; but the answer is just a suggestion that will simply satisfy the need to have something workable to start with.

Here LED, perhaps not entirely in earnest, suggests another strategy - and a very plausible one at first sight. It is the strategy of limiting the focus, year by year. ("...each year one or two burning questions asked by teachers, parents, principals, would be identified. A nation-wide call would go forth...") That strategy is open to the probably fatal objection that most schools would be facing other crises than the Chosen One, and need help before their particular concern comes to the top of the queue. She goes on to suggest some specific answers, if the Chosen One strategy does not eventuate.

I strongly support virtually all of her suggestions here, e.g., (i) that ranking as well as grading is of audiences we are serving here (so integrated reports are essential); (ii) that consumers will have very limited tolerance for lengthy reports and so panels should produce digests (consumer reports) at the beginning of the fuller report; (iii) that the evaluation process should have strong roots in the population it is supposed to serve, so local discussions should be set up every year that lead to state and thence to national priorities (some needs assessment is possible, even if a single focus is not appropriate), and these discussions should include the utility and quality of prior reports from the panels; the same should be done at meetings of the producers and managers of production outfits.

I would add a few additional thoughts.

A. Hold on for the same process for all panels. This view comes from the mission; we need to have comparability in order to get support placed where it will pay off most for students. Differences in panel methods will make comparisons questionable. This approach is strongly supported by logistical considerations, viz. continuity in staff support will be greatly facilitated if staff used to one panel can easily transition to another as illness, promotion, and departure take their toll. A simple example is the 'critical mass' problem; basically go for a low figure that will suit all panels (perhaps half a dozen submissions). A better example is the use of a common consumer report format: examples of this might be proposed in pre-meeting discussions by panelists and staff, and doing so should be an immediate priority. These tries could then be passed around for criticism in the immediate future all the way down to the constituencies and revised in the light of feedback, instead of being left to the first meeting's agenda. They will benefit greatly from this cycle of development. And yes, it's crucial that the panel argue about drafts of the consumer report, because no matter how much discussion there is, many of them will feel it was misunderstood by whoever wrote it up, if the writing up is done after they leave. In fact, I would run the discussion in terms of what wording (rating, etc.) should go in what place on the report, rather than in some other terms that are to be converted later.

B. "Calibrate" the panelists, i.e., spend one hour at the front end of the first meeting rating two or three simulated submissions, designed to illustrate most of the problems that come up, including the distinction between P and E. This pays off many times over in avoiding repetitive treatment of similar problems. (On this issue, I don't think one can wholly separate the discussion of Promising practices from Exemplary practices, because the former is, after all, just a bear cub trying to grow up. Even though only a few of them may become big bears, it helps to see the process as a continuous one.)

C. Stress that cost is part of the problem for most school, teachers, parents, and students, and so cost must therefore be considered in the solution. This means that some basic cost analysis skills will have to be available in one or two members of staff and preferably one or two of the panel.

D. While volunteer reviewers cannot be expected to review fifty submissions, they can be expected to review three, and the use of IRT and MAUT methodology can now extract a reasonable overall ranking from an overlapping set of such reviews. By exercising severe constraints on the length of submissions (e.g., restricting them to three pages) we would ensure that panelists could review half a dozen, which will yield a very reliable ranking without significant loss of validity. It should be remembered that it is a logical truth that no technology can integrate ratings into rankings, so the comparisons, if needed eventually, must be part of the original task.

E. On the consensus issue. Here I depart slightly from LED. I do not think any consensus is necessary. We are not electing Supreme Court justices here: we are trying to provide guidance. If the panelists disagree radically, something which will in fact only happen occasionally, let's say so and give the reasons for both sides - and let the consumers sort it out. What's the worst that follows in such a case? Only that several alternatives are supported by some and disdained by others; very well, that makes if possible for schools to go whichever way their staff prefers. And the state or the district or OERI will have to do some further work to resolve the issue at a more local level, in case it can be resolved. We will all benefit by getting some more work done on providing success/failure data. Let's not imagine this electorate has never encountered dissent over best interventions: it's the name of all politics. Even NIH can't get expert agreement about the advisability of annual mammograms for the sub-50 group; the electorate benefits enormously from learning that it's a close call, and why. It's not to be expected that nature has arranged judges who can recognize them at all stages of history.

F. On the face-to-face vs. other alternatives for panel meetings. By all means look at the RAND study as LED recommends. But keep in mind that we are in early days here, and the alternatives have not been well developed, let alone fully evaluated. Think of the issues about Delphi methodology; it took decades before we got a balanced view of the approach and its Achilles heel. Not only are there many different ways to run panels using each of phone or email or fax or snail mail, but there is a fast-changing situation with respect to how many prospective panelists could happily handle each of these approaches. It may be that providing workshops in these methodologies, e.g., at AERA, would be cost effective; it is certainly the case that many people one could not get to come to Washington will be willing to take part in a 'panel in the home' approach. And it is certainly true that the money released from travel costs, if usable for honoraria, could add to the list of willing participants many extremely knowledgeable people.

G. To the list of questions in the OERI document, Pat Ross added one that she was especially interested in having addressed. What kind of evidence of success should we be requiring or requesting from those submitting proposals? There are two parts to the answer: one concerns the instrumentation required and the other concerns the evaluation design required. A full answer must wait for another occasion, but two comments will indicate its direction. I have recently reviewed much of the literature on what is referred to as authentic assessment (the instrumentation) and much of the traditional literature on the issue of demonstrating educational merit (the evaluation design). I find strong reasons to agree with the view amongst educational measurement professionals about the former and to disagree with them about the latter - and this is not just a matter of looking for support after deciding on an answer.

On the former issue - authentic assessment - it now seems clear that most of these half dozen approaches have something significant to offer, with two qualifications. (i) They only offer something for very limited purposes (for example, portfolios are good for teaching if not used alone, but almost useless for external, e.g., state, testing). (ii) They are often supported, not only for legitimate purposes but for more than those legitimate purposes, by arguments that show extreme ignorance about more traditional methods (for example, by the argument that multiple-choice items can only test recall or recognition skills).

The limits of authentic assessment are well illustrated by the disastrous failure of the California and U.K. attempts to switch over to what were supposed to be more authentic approaches, a failure due almost entirely to the huge extra time requirements these impose on teachers. The benefits are widely if misleadingly extolled: Shavelson's work on assessment in science education illustrates what can be done in a responsible and limited way. I would just add to this that there is still considerable room for extension of the assessment armamentarium, e.g., by the use of multiple rating items, a type of test that can be scored mechanically but that tests higher-order skills very directly (I outlined this approach at the NCES national meeting on assessing critical thinking skills).

On the latter issue - what it takes to show success - I fall in with the great applied statistician Frederic Mosteller, because I feel that his more general perspective is more relevant than the local (i.e., education-centered) perspective of the usual authorities. Those in education are congenitally committed to dealing with very small changes in the pre/post test scores and develop very subtle ways to do so. But small effects of this kind are tricky to pin down because of regression effects and other artifacts. Mosteller's view, expressed by his introduction of the term "interocular differences", is that one should only use statistical significance as a guide to one's efforts to find something more valuable, i.e., differences that hit one between the eyes, hence are 'interocular'.

I take this approach to education; if one can't produce interocular difference between the pre and post test scores, using ABBA designs and externally validated test items, one should give up teaching. So I would recommend sticking with commonsense on this point: expect those who seek recognition, whether it be as promising or exemplary, to submit data showing massive learning gains on tests that are independently validated as testing an important slice of the important relevant skills.

VOLUME I ***Section II. Papers Examining other Review Activities to Identify the Best***	
A Synthesis and Integration of U.S. Evaluation Efforts to Identify Promising and Exemplary Educational Programs, Products and Practices -- Gary Borich, Univ. of Texas, Austin (July 1997)	**E**
Examining Federal Approaches Outside the Department of Education to Identify and Disseminate the Best -- Karen Bogart, Anne Steinmann Institute (June 1997)	**F**
Notes on the Federal Dissemination Experts Meeting: "Advising the Department of Education's Office of Educational Research and Improvement on its Design of a Systematic Consumer-oriented Evaluation System to Designate Promising and Exemplary Educational Products, Programs and Practices" Oct. 12, 1995, Academy for Educational Development, Washington, DC. -- Karen Bogart, Ann Steinmann Institute	**G**
Standards and their Use in the Food and Drug Administration (FDA) -- Gerald (Jake) Barkdoll (Jan. 1996)	**H**
Review of Foundation, Associations, and Non-profits Practices in Designating Promising and Exemplary Programs-- Janet Carter and Diane Schilder (July 1997)	**I**
Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas -- John Luczak & Joan Ruskus, SRI International (July 1997)	**J**
A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science -- Carol Muscara, Computer Technology Services, Inc., Rockville, MD (Nov. 1996)	**K**

E

A Synthesis and Integration of U.S. Evaluation Efforts to Identify Promising and Exemplary Educational Programs, Products and Practices.

Gary B. Borich
The University of Texas at Austin

This report syntheses and integrates some of the ideas and concepts presented in four OERI commissioned reviews of U.S. evaluation efforts to identify promising and exemplary programs, products and practices. The agencies whose evaluation efforts were reviewed consisted of educational organizations outside the government (Luczak and Ruskus, 1997), government agencies outside the Department of Education (Bogart, 1997), math and science programs (Muscara, 1996), and foundations and nonprofit agencies (Carter and Schilder, 1997). This report synthesizes the distinctions among the four review efforts, integrates their common elements and identifies trends and patterns that hold promise for informing the FindBest system.

The four reviews were commissioned without a restrictive framework in order to allow the authors to establish a format responsive to the mission and goals of the agencies being reviewed. In this regard each set of authors established, with guidance from OERI, their own template for organizing their thoughts, choosing the lenses through which they would see and record data and articulate their findings. Therefore, the variety with which the authors chose to organize and format their reviews left this author with many of the same challenges as were so admirably met by the reviewers themselves. To capture this variety, this report provides a high altitude aerial view of what these reviews have to say to OERI efforts to distinguish promising and exemplary programs, products and practices. Accordingly, this work is organized around (a) types of resources reviewed, (b) types of information acquired by the reviewers, (c) types of review efforts, (d) types of criteria of program, product or practice effectiveness, and (e) the relationship of what was learned to the OERI expert panels.

Types of Resources Reviewed

The resources reviewed by the four commissioned authors included programs, products, practices, procedures, policies and demonstrations, here after referred to as *resources*. By and large, the most common resources were programs and products, with some focus on practices. The reader is referred to the original reviews for examples of each, but for the purposes of this review the distinctions among these labels and their use across agencies outside government, government agencies outside the Department of Education, math and science programs, and foundations and non-profit agencies (hereafter, called *domains*) were of primary interest.

Programs. The label "program" was commonly applied to complex interventions representing multiple components. An example would be a program to raise math achievement of elementary learners, which also had a component to train teachers to implement the program and another to teach parents strategies to help their child benefit from the program. From a research or evaluation, these omnibus agents of change represented bundles of independent variables and multivariate indices of "program" effects. They represented a particular challenge to some agencies in weighting the various components within a program for which evaluation data in form and quality differed markedly. Other agencies seemed to have ignored the component parts in favor of the holistic nature of the program and those outcomes most consistent with the goals of the agency. In general, a holistic view of programs, focusing on their global effects across rather than within components, was the method of choice for evaluating this type of resource.

Products. Products were simpler to define than programs and shared many of the same common ingredients across agencies. These common ingredients included (a) easy transportability from site to site and sometimes from one context to another, (b) self-containment, requiring few if any external resources that were not already available in the context in which it was being implemented, (c) modifiability to other contexts and populations, and (d) applicability to quantitative evaluation which sometimes included quasi-experimental designs or some measure of control representing the implementation of the product across sites for comparison.

Practices and Procedures. Practices and procedures were, for all intents and purposes, indistinguishable. What one agency might refer to as a practice another would refer to as a procedure. In either instance this category of resource was the most difficult to define as an intervention or change agent and its outcomes tended to be least targeted. An example of a practice would be instituting a prenatal examination for at-risk mothers, or the implementation of a "procedure" that identifies initiatives for youth employment. Even though a few agencies might choose to label such resources "programs," these types of resources shared little similarity to those which were labeled programs by most agencies and were seldom different from those resources labeled "procedures."

Policies. Policy in the context of the reviews often referred to administrative arrangements that allowed a host agency to achieve its goals more efficiently. Although not many agencies reviewed "policies," some used the label interchangeably with practices and procedures. The distinctive feature of a review of policy, however, was that the policy was integral to the institutional structure and management system within a particular setting and sometimes indistinguishable from it. While procedures and practices sometimes were discussed in the context of materials or other tangible resources that helped the procedure or practice become implemented, such as a "handbook" for human service delivery personnel to implement a drug rehabilitation procedure, or a "checklist" for a special educator to determine if a particular home intervention practice had been implemented, policies were often discussed in the context of systemic concerns from the view of management. Policies, in other words, were seen as directives for the organization that left little observable mark that would be considered independent and evaluatable by themselves. Perhaps for this reason policies were not often the subject of review and, when reviewed, provided some of the weakest data for unambiguously determining their merit.

Information Acquired by Reviewers

The information acquired and subsequently reviewed by agencies included:

Type of Resource. These data provided some reference to whether the resource was a program, product, practice, procedure, policy or demonstration. But, more importantly, the authors presented under this or similar heading (e.g. "background") the prerequisites, if any, that the resource had to possess in order to enter the review process. For example, in some instances the resource had to have gone through a previous review by an agency, such as State Department of Education, that may have accorded the resource some distinction, or had to have met State and/or Federal compliance requirements.

The "criteria to enter" the review process varied considerably across domains with many agencies welcoming voluntary submissions after being sent information describing the mission of the agency and some details of the submission process. In these instances, the criteria to enter reduced to the resource fitting the mission and immediate goals of the agency.

Population. Population data were considered by the reviewers as the audience or audiences the resource was to serve. Needless to say, the range of populations was considerable given the breadth of the four domains of agencies reviewed, which included non profit agencies and foundations many of whom focused on adult or special populations, math and science resources which customarily dealt with the public schools, and agencies both in the government (excluding the Department of Education) and outside, such as general public service agencies, such as the Federal Emergency Management Agency (FEMA), whose goals were established on need without regard to population. However, these varied foci did impact the type of resource reviewed by an agency. In the case of FEMA, programs and products, which made the best fit to that agency's mission, were not as prominent as practices and procedures,

The target population for an agency seemed to be one of the strongest forces influencing an agency to consider a certain type of resource, such as "program for elementary school math" or "practice for prenatal care of adolescents." Concomitant with this was that some audiences were only found within certain institutional contexts (e.g. the public schools) which had administrative structures (e.g. math and science departments) whose work tended to be defined by some types of resources (e.g. programs) more than others (e.g. procedures or demonstrations). Hence, populations and resources were interrelated, with the audience or population sometimes focusing the type of resource that came to the attention of the agency for possible review. It should be noted that institutional conventions, administrative boundaries and categorical labels, that come to define populations, in some instances, may have had as much to do with deciding for an agency what resources would be reviewed as did mission and goals.

Evaluation Methods. The methods used by agencies to screen and evaluate the submissions they received or solicited could be grouped into five categories.

Relying on key staff. One approach used by almost every agency was to rely on experienced and knowledgeable staff to screen initial entries, who later may participate in a more thorough review of the resource. The first step in this process was to complete an initial comparison between the goals of the agency and the outcomes claimed for the resource. Of particular note during this stage was the delivery system by which the resource intended to reach its audience and achieve its outcomes. Although the details of this process were not always apparent from the available data, it appears as though a large portion of submissions were rejected by most agencies at this level due to (a) the inapplicability of the resource to the agency's mission and (b) the inexactness in which the expected outcome(s) were stated, leading the agency to cast doubt on its suitability for further review. Most agencies had a screening and review process in place, if not to evaluate the resource, then to delimit the number of resources needing further review in order not to overwhelm a small support staff. After an initial screening the agency may have used one or more of the following evaluation methods to complete a more thorough review.

Independent external reviewers. One approach frequently used by the agencies was to commission external reviews to examine the resource and its claims of effectiveness. Depending on the resource, this might involve actually physically examining the resource or in rare instances, due to cost considerations, going on-site to observe its implementation and/or operation. In most other instances, however, the independently contracted reviews, which often involved more than a single reviewer, were completed from "paper" documentation and descriptions submitted by the developer. In this mode of review, the reviewers seemed not to have contact with one another during the review process or know that other reviews of the resource were being completed simultaneously. The work of collating and integrating the reviews, therefore, fell upon key agency staff, who would make a final decision from the reports of the reviewers as to the merit or distinction accorded the resource. This approach may have been combined with one or more of the following evaluation methods.

Outside panel review. This approach convened a panel of reviewers from outside the agency to collectively consider the merit of the resource, paralleling the OERI expert panels. The structure provided to the panels seems to have varied considerably, from leaving the panel decide its own procedures to following a structured list of checkpoints provided by the agency.

The amount of structure provided the panels appeared to have been a crucial issue. In most cases it appeared the agency knew how it wanted the panel to organize its deliberations, i.e. ordered and provided the agenda, provided checklists or other rating devices, and provided a clear statement of the criteria against which the resource would be compared. In some cases, this agenda seems to have left little room for more global comments and additional qualitative evidence, such as interviews with users, anecdotal evidence of "satisfied" consumers, and participant testimony which might have afforded the panel more independence to make decisions outside the framework established by the agency. This tended to be the case more with respect to resources reviewed by educational agencies outside the government and in math and science programs than review efforts in government agencies and non profit agencies and foundations. With the latter, there appeared to be less of a trend to emulate a "hard data" model for the review of the programs and more flexibility to accept review efforts involving more qualitative and/or non quantifiable data, such as subjectively determined congruence of the goals and effects of the resource with the mission and audience of the agency.

Government agencies and non profit agencies and foundations seemed to share the flexibility to seek out and use in the review process more holistic, qualitative data as well as hard to document professional experience, insight, and intuition. And, while these data were not used to the exclusion of harder data in the form of ratings and outcome data from the field, they did seem to recognize the integration of the quantitative and qualitative more than agencies outside the government and math and science programs.

It also may be that government agencies and non-profit agencies and foundations had the freedom to operate in a less "entrepreneurial" spirit than did math and science programs and educational agencies outside of the government, whose audiences may have sought the sophistication of hard data, such as the case of achievement data normally collected and required by math and science programs. Non-profits and government agencies, although no less accountable to the public in general, may have been more buffered from direct competition among consumers. In other words, goals, philosophies and missions may have played a greater role in guiding the determination of merit of resources by non-profits and government agencies, while the need to be accountable to specific audiences may have influenced math and science program reviews and the reviews of resources by non governmental educational agencies, where the case may be made that the link with specific consumers is more direct.

Although infrequent due to cost, site visits by the panel members to examine the implementation and/or operation of a resource were occasionally employed within this method of review.

Contracted organization. A fourth approach to the evaluation of resources was to outsource the evaluation process to an agency or vendor skilled and experienced in doing program and product reviews, who may themselves be program developers. This approach was used more by non-profit agencies and foundations than by review efforts in the other domains. Non-profits and foundations seemed to have more flexibility both in how funds were gathered for the operation of the agency or foundation and in how it was spent, allowing more options for the review process. The expense and allocation of this responsibility to others outside the agency seemed to make this an unattractive option for most of the other domains, whose funds for this activity would necessary compete with other initiatives. It is also likely that staff that could manage and/or conduct reviews in-house were in greater supply in some of the other domains than in non-profit agencies and foundations whose activities might be more limited to gaining sponsorship and promoting specific missions and directives. From the evidence available, this option provided a satisfactory alternative for some agencies, but was clearly avoided by others who wished not to delegate decisions and criteria of merit to an external contractor, limiting their control over and management of the evaluation and review process.

Conduct a demonstration. Although infrequently reported, a few agencies either commissioned a demonstration of a promising program idea or commissioned an evaluation of one that was already being "demonstrated" through the auspices of another agency. Cost considerations made this a more feasible alternative among more flexibly budgeted non-profits than might be expected across the other domains. However, disregarding cost, this alternative provided the opportunity to achieve some of the strongest field data among resources, most often in the form of experimental or quasi-experimental evaluation data. The idea of specifically funding a demonstration or using a demonstration project funded by another agency for evaluation purposes provided a measure of control and comparison unknown to many other data collection formats. These were instances in which a resource had been determined by the agency to have strong ideational merit and/or appeal as well as the promise of providing a direct response to an extant problem of some public concern. In these instances, where the resource was a response to an acknowledged public concern, and, therefore, there was the need to have the quality of the evaluation data unassailable, a demonstration and evaluation was funded to provide unambiguous data for determining further funding, development or implementation of the resource.

Criteria Of Effectiveness

Criteria for determining the merit and effectiveness of a resource could be divided into two broad categories: (1) less formal evaluation data used by the reviewing agency and (2) more formal data submitted by the applicants themselves as justification for a designation of merit or distinction.

Less Formal Effectiveness Data Used By the Reviewing Agency. This category represented less formal data that reflected some of the standards used to judge the merit of a resource. Although not all agencies invoked all of the following, most used several, if not many in initially screening an application for a distinction of merit.

Currency and timeliness. Many agencies either explicitly or implicitly required the resource to serve some extant need salient to the constituent groups they were serving. In other words, there were indications that most if not all agencies had written or unwritten priority areas in which resources were known to be needed. These priority areas sometimes were clearly communicated to potential applicants and, in some cases, they were not. The latter become evident by the number of submissions that appeared to be rejected at this first level for being out of the purview of the agencies perceived mandate or not fulfilling high profile needs within it. In this case, the question asked of the resource was "did it respond to a documented need known to exist for one or more client groups being served by the agency." It was often unclear what evidence an agency used, however, to determine extant needs other than identifying "high profile" concerns among client groups. In some cases these needs may have come from professional judgment, insight, and intuition within the agency and at other times may have resulted from more or less formal needs assessments conducted or commissioned by the agency and/or documentation in the popular and professional literature.

Correlation with standards or state and national frameworks. When applicable most resources appeared to be "screened" for their adherence to state and federal legislation and policies and findings of professional groups. Special education programs, for example, were a prime example of this level of scrutiny in order to determine if they met state compliance standards. In other instances, agencies were "tuned in" to policy, ethics and advisement's of government and professional groups who may have had documented standards for a particular type of resource in their area. As with many of these "less formal" criteria, the judgment made by the agency could be described as qualitative. Little hard data were available to help with the judgment and most of the decision making appeared to be at the level of "screening" the resource more from the view of violating standards and norms, then from the requirement of actively contributing to them. In other words, this criterion, as with some of the others below, tended to be measured in a "yes/no" checklist fashion as to whether it was congruent with state or professional standards based on the judgment, experience and knowledge of agency staff.

Equity, lack of bias. This criterion appeared to be less subjective than the former two. Although hard data as to a resource's "equity" or lack of bias seemed rare, this standard for evaluation appeared more targeted and uniform across agencies. This may have been the result of the attention this standard has received among professional groups and the many extant examples of bias that have been reported in the popular and professional literature. In other words, it seemed as though this criterion was promoted more uniformly across agencies and that agencies seemed to have a clear grasp of what would not be acceptable. This impression was not conveyed in how the former two criteria--timeliness and correlation with standards--

were implemented, in which the emphasis across agencies appeared to be more uneven and degree of adherence more subjective. In this instance, agency staff seemed to look proactively with clearly defined issues of equity, as opposed to examining the resource for something that might "pop out" to tell them it wasn't timely or violated some government legislation or professional policy. Thus, this less formal criterion was among the most consistently applied.

User appeal, user involvement. Although related to currency and timeliness, some agencies seemed to place special importance on the potential "audience" appeal of a resource and the involvement, potential or documented, of participants and agencies in integrating it into its functions and services. This criterion seemed to reduce to "ease of use," as the prime objective, which, in turn, was perceived to determine user appeal and involvement. The fact that a submission application documented a history of use and participation, particularly with the agencies expected to use the resource, counted very much for the resource. Some included a history of the resource over many years of successful adoption and the number of participants that have used or have been served by the resource.

Usefulness to others, in another setting or content area. Some agencies not only expected a history of use, but also that the resource be adaptable to other areas or settings consistent with the agency's mission and goals. The question seemed to be "how much mileage" could the agency get out of the resource, or achieve the "biggest bang for its buck." This may have involved how tractable the resource was to serving broader audiences than its designers may have originally intended. The generalizability of a resource, for example, a math and science program from the secondary to the junior high level, would be an example. Some agencies seemed to have looked for generalizability or adaptability across populations, school levels, or content area. For example, could a procedure designed to attract teenagers to seek prenatal exams also be used or adapted to attract a more general population of women to breast exams, or could a program for the treatment of drug abuse also be used by community mental health centers to serve an alcohol rehabilitation program? The less the resource would need to be adapted to new populations, educational levels, or content areas, expending resources of the agency, the more it could be expected to achieve this criterion.

Cost effectiveness. In the public as well as non-profit sector, cost versus effectiveness was a prominent concern. Although difficult to quantify with the data submitted to agencies, many agencies, nevertheless, appeared to place importance on this criterion. The submission of cost data, therefore, was particularly important for a resource to achieve this criterion, and where possible, its relationship to effectiveness shown, although this connection was seldom made explicit. In other words, costs and effectiveness were usually two distinct classifications of data available to the agencies and their relationship would often have to be inferred. Nevertheless, costs, particularly cost of implementing the resource to reach its target group, was an ingredient in determining the degree to which a resource achieved this criterion. Effectiveness data appeared to be the counterpoint against which costs could then be compared, based on the experience of the agency staff or commissioned reviewers.

Built-in user support. Although a subtle and less often cited criterion, this characteristic of a resource, when invoked, seemed to count in favor of a resource. For some resources the criterion was inapplicable, when, for example, a school-based science program was almost entirely self contained and uncomplicated enough to be implemented and/or serviced by on-site personnel. But, in other cases, for example with a computer software component utilized in one FEMA's emergency management programs, built-in user support was essential to the functioning of the larger program. Therefore, while not all or even most resources might need to achieve this criterion, for others it appeared to be a criterion that was essential.

Resource organization and management. This criterion appeared to emphasize the efficiency with which the resource could be operated and managed. In other words, what material and human commitment did it take to implement the resource in the real world? Resource submissions, therefore, had to document successful implementations in order to demonstrate this criterion. The most successful resources in doing this would point to demonstration sites in which the resource was "up and running" and therefore could be observed, participants questioned and management interviewed as to its smooth functioning within a particular institutional or social context. The emphasis seemed to be on the capability of the resource to be run successfully in the real world, so the life-like quality of a demonstration site or examples of successful implementations were important. A distinction that appeared to be important was whether the record of implementation was particularly colored by "in house," "pilot-test," and "trial," runs that were contrived for the convenience of the developer as opposed to implementation and successful demonstrations of the resource in the real world context in which it was expected to operate.

Collaboration with other agencies. This criterion viewed a resource more favorably if it could successfully achieve its goals in collaboration with other agencies. Sometimes this was assumed to mean that the resource could expand its goals to those greater than originally intended by collaborating with other agencies. This criterion, therefore, would be most appropriate for resources whose mission could be shared with other resources. For example, in one case involving a results-based accountability project, the Pew Charitable Trusts and the Ford Foundation both contributed simultaneously because they shared missions consistent with the objectives of the resource. Or, when the National Coalition of Advanced Technology Centers (NCATC) and The Center for Occupational Research and Development (CORD) shared a mission to develop Advanced Technology Centers their collaboration resulted in the development of a new resource. Where such collaborations were reported and documented, the resource achieved this criterion.

Outcomes consistent with agency agenda and mission. The most criterial attribute for a distinction of merit was the resource's compatibility with the agencies agenda and mission. This criterion, understandably, was the most important in rejecting submissions without further review. Therefore, it was customarily completed before any other criterion of merit was applied, informal or formal. This criterion was different, however, from that which examined the extent to which the outcomes of a resource appeared consistent with federal, state, and professional standards, as discussed above. Here, the standard was internal to the agency, determined by its constituent groups, its own legislative authorization, sponsors and board of directors. This standard in the form of mission, goals or agenda appeared to be better documented in some domains than others and better documented across some agencies within a domain than others.

Where an agency's mission, goals or agenda were so broad and general as not to provide practical guidance for reviewing a resource, this criterion of merit appeared of lesser relevance. Non-profit agencies and foundations appeared to have the most articulated and specific missions. Math and science programs appeared also to have well articulated goals, often in the form of behaviorally stated objectives, with federal agencies outside the Department of Education appearing to exhibit less specific missions and goals, presumably because of the broad audiences they were mandated to serve. In other words, it would be easier for a broadly defined resource, in terms of population, level of schooling, and content area to achieve this criterion in federal agencies outside the Department of Education than it would for this same broadly defined resource to achieve this criterion in the domains of non-profit agencies and foundations, and math and science programs. The fourth domain, educational organizations outside the government, appeared so diverse that no consistent trend concerning this criterion could be found. Nevertheless, the capability of a resource to harmoniously blend with and contribute to articulated goals of an agency appeared preeminent in the review process.

More Formal Effectiveness Data Used by the Reviewing Agency. The more formal methods of determining the effectiveness of a resource included quantitative and qualitative data, data derived from self studies and site visits. The procedures of many agencies included combinations of these in addition to the less formal methods identified above. Typically, but not always, formal evaluation data provided information about the audience, criteria of determining effectiveness, process used in conducting the evaluation and the results.

Quantitative and qualitative criteria. These data consisted of combinations of descriptive statistics (usually in the form of ratings), correlation data, and occasionally quasi-experimental data. From the information available, little is known about the details of the studies themselves, but some indications of how quantitative and qualitative data were collected were available and, in some other cases, could be inferred. Few, experimental or quasi-experimental studies involving variable manipulation and control were noted.

The most frequent use of quantitative criteria for determining the effectiveness of a resource were ratings and checklists, either completed by reviewers singly or as a panel commissioned by the agency, or by the submitter indicating their own or commissioned results of an evaluation. Typically, a five-point Likert scale and/or checklist was used to record perceptions of the resource in the context of questionnaires and surveys in the case of users, or in the context of agency rating scales in the case of commissioned reviewers or agency staff. In essence three types of data sources were utilized: reviewer, resource users, and agency staff.

Reviewers and resource users tended to use a general or holistic approach to rating the resource. Agency staff, however, sometimes had the additional responsibility to rate a resource and weight its component parts from information submitted on an application form. These forms appeared to differ considerably across agencies, but some general procedures could be identified. For example, in the case of the National Youth Employment Coalition, the application form asked the resource to respond to questions in each of several categorical areas, such as (1) Purpose and Activities, (2) Organization and Management, (3) Youth Development, (4) Workforce Development, and (5) Evidence of Success. Each was accorded weights in the form of points that would be used by agency staff to evaluate the responses to questions posed on the application form in each of these areas. For example, "Purpose and Activities" was assigned 0 points, but was essential to agency staff evaluating all other sections and assigning a final score. The other four categories were assigned 50 points each for a possible total of 200 points. Each category had from five to eleven questions that had to be answered by the applicant which agency staff assigned up to 50 points.

Other programs and agencies used similar methods. For example, the Special Education Exemplary Programs Review of the Idaho Public Schools divided their application form into seven categories, "Student Outcomes," "Least Restrictive Environment," "Collaboration," "Instruction," "Parent/Community Involvement," "Personnel," and "Resources." Unlike the National Youth Employment Commission highlighted above, this program review weighted each of the seven categories differently (from 10 to 35 possible points), with an extra 5 points assigned to a submission that indicated one or more "unique features" in any of the categories for 185 possible points. Still other reviews used Likert scales that were associated with specific questions within categories to determine if the program under review met various criteria.

With the exception of guiding questions further delineating what was to be rated, gradations between points on the scales were subjective, resting on the experience and expertise of the rater. In other words, the criteria for assigning a "4" versus a "3" or "3" versus "2" on the scale was not objectively apparent. Similar scales were used by panel reviewers outside the agency and, sometimes, in evaluations conducted or commissioned by the developer. However, this process differed from agency methods in that it did not usually have preassigned weights to the items being evaluated and, instead, tended to average numerical ratings within and across scales to arrive at a total evaluation, say, between 1, "unacceptable" to 5, "exemplary."

In all but a few evaluations there was little attempt to place resources distinguished with high ratings into categories, such as "exemplary" and "promising." Acceptable and unacceptable, winners and losers, selected and discarded could have been some of the labels that represented the results of these decisions. This is not to say that a two tiered system, such as promising and exemplary, would not have been used if standards for their execution could be articulated, but rather that few such distinctions were encountered in practice. It seemed clear that if such a distinction were to be made, the terms used to make the distinction, e.g. promising and exemplary, would have to be amply documented with standards, rules of evidence and a decision-making process that would resonate with the specific goals and agendas of the agency. In other words, the agency would have to see a practical need to make a two (or more) tiered designation of merit and to provide differential rewards across categories. Most agencies had not thought about any finer gradations that might separate resources or the different rewards that would accrue to them, for example, labeling some resources "promising," but not "exemplary" to encourage them to apply again in a more mature state or to provide the resource some tangible reward in terms of dissemination, publicly stated acclaim, or even funds for further development to reach "exemplary" status.

Another pattern with regard to this criterion was the intensive use of human insight, experience, and intuition in making distinctions of merit. It was as if some of the agencies were saying that the exclusive use of statistics may show the performance of a resource in light of the average but may not provide a complete picture of reality. In other words, the intensive use of review panels and the application of professional experience, insight, and intuition in actually interpreting, weighting and integrating quantitative data with the qualitative seemed to be a way of capturing the irregularity in the performance of a resource, even though its average performance may be impressive compared to its competitors. These review efforts seemed also to provide the flexibility to sometimes weigh a resource's "best" performance more than its average performance.

Finally, it needs to be noted that a number of review efforts used both qualitative and quantitative data, with an important caveat. It was not their intent to integrate these different data bases by requiring qualitative methods to be more quantitative with, say, numerical scales and ratings, or quantitative methods more qualitative with, say, interviews, in an effort to uniformly merge or blend findings from the two. Instead, at least some of the agencies seemed to acknowledged the central point that decisions about resources did not so much require the methodological blending of qualitative and quantitative data, as some proponents of numerical aggregation suggest, but rather the convergence of their separate and distinct results. Hence, a resource that received lower evaluations from more formal data (e.g. quantitative) would not necessarily color the interpretation of the resource as it may have been described from another less formal source of data (e.g. user appeal, timeliness, equity, ease-of-use, etc.). This may have been one reason for the preponderance of agency staff and review panels who did not simply count, add up and aggregate the evidence but played an active role in placing judgments upon the perceived truthfulness of the data to inform them of the performance of a resource against the backdrop of the goals and agenda of a specific agency and its constituents.

Data derived from self-study. In some cases the application process required by an agency encouraged a self-study by the resource. In the best cases, these applications required the resource to report on their own self-evaluation and to do so with objectivity. This was sometimes accomplished by the specificity of questions posed in the application procedure and the detail in which the applicant was expected to respond--if only by the length and breadth of questions, and space provided for responding. It also was accomplished through the judicious selection of questions, some of which asked the applicant to state the claims being made for the resource and provide objective (for example, third party) evidence that the claims have been met. Some questions evoked responses that required or encouraged the applicant to state the relative advantage of the resource over realistic competitors, a history of those who are or had been using the resource, evidence of its compatibility with the larger organizational units, systems, or programs of which it would be expected to be part, and its costs and payoff (cost-effectiveness). In other words, depending on how carefully the submission application was prepared and worded, it more or less served as a self-study--in some cases invoking a responsiveness to evidence and accountability that otherwise would not have been undertaken by the resource. All things considered, a self-study, encouraged by a carefully thought out application process, appeared to be one of the most useful and practical means of determining the merit of a resource.

Site visits. Although infrequent, site visits appeared to be a tool for corroborating both less formal data, and other types of formal evaluation data. Site visits have long been a tool for evaluation, particularly program evaluation. The American Psychological Association, as well as the National Council for the Accreditation of Colleges of Teacher Education routinely charter university training programs from the results of site visits. Although site visits would not be applicable to the majority of resources, particularly those in the form of products that could be examined outside the context of an instructional structure or system, the evaluation of other "high profile" programs would be deserving of this approach. It's use, however, did not appear to be consistent or frequent for the resources reviewed across all four domains. In those cases in which a site visit was employed it appeared to be of particular consequence in the overall evaluation of a resource and the data, insights, and judgments derived from it accorded considerable weight. It's main value, however, seems to have been the focus it provided on specific criteria through the formulation of a template of desired standards and criteria to guide the observation (site visit) process and the extent to which it could examine the resource in the on-going context of an institution whose relationship with the resource could enhance or diminish its effectiveness. When conducted, the site visit appears to have illuminated, enriched, and expanded data from other sources, both more and less formal.

Relationship of What Was Learned to the Expert Panels

The above review raises several questions relevant to the conduct of the Expert Panels. Among these are (1) With what criteria can the expert panels uniformly evaluate a resource for distinction? (2) How might a two tiered system of promising and exemplary resources be "sold" and justified? (3) How might potential submissions be directed and managed to provide the best and most relevant data to the Expert Panels for review? (4) How might consumers participate in the verification of the effectiveness of a resource? and (5) How might publishers and developers be encouraged to participate in the FindBest system? Some of the above questions are stimulated by what was found across reviews in the four domains, while others were suggested by what was notably absent from the reviews.

1. With What Criteria Can the Expert Panels Uniformly Evaluate a Resource for Distinction?

From these reviews a robust set of effectiveness criteria emerged. These criteria included:

- ☐ The currency and timeliness of a resource,
- ☐ it's possible correlation with standards implied by state and national legislation and policy,
- ☐ the resource's equity and lack of bias,
- ☐ its user appeal and involvement,
- ☐ its usefulness to other populations, other settings or other content areas,
- ☐ its effectiveness versus anticipated costs,
- ☐ the degree of built-in user support,
- ☐ evidence that the resource has been operated and managed in natural contexts,
- ☐ the degree the resource encourages collaboration with other resources or agencies, and
- ☐ the extent to which the outcomes of the resource are consistent with the goals and mission of the department, agency or program it is to support.

Responses in the form of a five-point scale for each of these criteria completed by the expert panel members from application documentation submitted by the resource could serve as one component of the review process.

2. How Might a Two-Tiered System of Promising and Exemplary Be "Sold" and Justified?

These reviews indicated little evidence of or support for a two-tiered system of distinction. Since such a system would be unfamiliar to those likely to participate in the FindBest system, some incentives to submit resources for review under a two-tiered system would seem in order. One alternative would be to set aside funds for resources designated as "promising" for the purpose of further development, evaluation and dissemination. The incentive for participation would be that, if chosen as promising, the resource could compete for funds that would be specifically allocated to assisting promising programs to achieve exemplary status in a subsequent review. Although funding would be competitive among resources achieving promising status, agencies might be encouraged to provide stronger evidence of development, evaluation, and dissemination to achieve this status. In addition, from the promising resources that are funded, other resources would see what kinds of evidence of further development, evaluation and dissemination needs to be sought to achieve exemplary status and, therefore, provide a model for other resources. Likewise, funds could be designated for resources designated promising for the purpose of further development in other target groups, e.g. at-risk, and/or other levels of schooling or content areas that might bring it closer to or expand an agency's mission and goals.

3. How Might Potential Submissions Be Directed and Managed to Provide the Best and Most Relevant Data for Review By the Expert Panels?

This question concerns how the submission process might encourage the most relevant data set for the evaluation of a resource. One approach for accomplishing this would be to design a submission process that would also require the resource to complete a "self-study" prior to the submission process. As an example, the requirements for submission might mirror most or all of the criteria of effectiveness listed in Question 1. In other words, each submission would be asked to submit data for each criterion listed vis-a-vis currency and timeliness, correlation with state and national standards, equity, user appeal and involvement, etc. Their responses according to each criterion would then be rated by each panel member and discussed collectively to provide a common framework with which to begin panel discussions of a particular resource. It would not be assumed that these criteria and their rating would constitute the sole evaluation of a resource, but would provide a basis for substantive comparisons and discussion among resources within the same panel and encouragement and direction to the resource for undertaking a self-examination in those areas which could lead to more thorough evidence for evaluation and review by the panels. At the end of a self-study, resources would have a clearer notion of whether submission is warranted and would have been provided some guideposts as to what types of evidence to gather to strengthen their own evaluation process. From this semi-structured submission process could be derived example "data sets" in the form of profiles of evaluation evidence that could be used to illustrate the gradation from promising to exemplary that could be shared in informational brochures describing the submission process.

4. How Might Consumers Participate in the Verification of the Effectiveness of a Resource?

The above reviews indicated that consumer participation in the evaluation of a resource was inconsistent within domains and even more varied across domains. One way of encouraging consumer participation in the review process is to require consumer reactions as evidence of effectiveness for promising and exemplary status. A distinction is made, however, between participant responses to a program in a quasi-experimental, simulated or pilot-test context and consumer reactions in a naturalistic setting. Since many evaluations tend to be summative, consumer reactions in a naturalistic setting should serve the formative purpose by identifying areas of desired revision and modification to a resource. Therefore, consumer reactions to a resource, as it is being implemented in the field, should be among the criteria required for achieving promising or exemplary status, and therefore added to the list of criteria in Question 1 and required for the self-study suggested above.

Furthermore, data should be targeted to identifying areas of perceived strengths and weaknesses of a program--as opposed to its overall effectiveness--and go beyond statistical indices that may mask variation in responses to a program by specific subgroups of consumers. Qualitative indices of consumer satisfaction, case studies, interviews and vignettes of participant dialogue that capture consumer confidence would serve this formative purpose and provide natural language benchmarks for the consumer verification of a resource.

5. How Might Publishers and Developers Be Encouraged to Participate in the FindBest System.

From the above reviews, there appeared to be few systematic efforts to uniformly request submissions from publishers and developers who might benefit from having a resource distinguished. One incentive for publishers and developers would be for the Department of Education to formally communicate the promising or exemplary status of distinguished programs and products to potential adopters. The manner in which promising and exemplary programs and products are communicated to the public could include the following:

1. A Department of Education publication specifically designed to identify and describe promising and exemplary programs and products that would be made available to the public. Potential users could subscribe to a nationwide mailing list to receive the publication.

2. Department of Education notification to the professional associations to which the program or product may have relevance, indicating its promising or exemplary status. Associations frequently provide awards, ceremonies and recognitions to projects in their field that have attained distinction. Announcements at professional meetings and in newsletters are customary channels of communication to association members.

3. A Department of Education Web page on the Internet indicating the programs and products achieving promising and exemplary status. Entries could be organized and cross-referenced by topic areas, audiences

and levels (for example, populations, grades, levels of schooling, institutions, etc.) to individualize access for the user. With each entry, a description and the source of purchase could be provided to create the "sales" or user incentive for projects to apply for promising or exemplary status.

Reports

Bogart, K. *Examining Federal Approaches Outside the Department of Education to Identify and Disseminate the Best.* Office of Educational Research and Improvement Commissioned Report, U.S. Department of Education, U.S. Department of Education, 1997.

Carter, J. and Schilder, D. *Review of Foundation, Associations and Non-profits Practices in Designating Promising and Exemplary Programs.* Office of Educational Research and Improvement Commissioned Report, U.S. Department of Education, 1997.

Luczak, J. and Ruskus, J. *Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas.* Office of Educational Research and Improvement Commissioned Report, U.S. Department of Education, Washington, D.C., 1997

Muscara, C. A *discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science.* Office of Educational Research and Improvement Commissioned Report, U.S. Department of Education, Washington, D.C., 1996.

F

EXAMINING FEDERAL APPROACHES OUTSIDE THE DEPARTMENT OF EDUCATION TO IDENTIFY AND DISSEMINATE THE BEST

Karen Bogart, Ph.D.

May 1997

Commissioned Paper

Prepared for

The Office of Educational Research and Improvement
555 New Jersey Avenue
Washington, DC

Examining Federal Approaches Outside the Department of Education to Identify and Disseminate the Best

TABLE OF CONTENTS

INTRODUCTION

METHOD

OFFICE OF ADOLESCENT PREGNANCY PROGRAMS

THE FEDERAL EMERGENCY MANAGEMENT AGENCY

THE SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION

AGENCY FOR HEALTH CARE POLICY AND RESEARCH

GENERAL ACCOUNTING OFFICE

HEALTH CARE GUIDELINES (HEDIS): SAMHSA AND AHCPR

NATIONAL ENDOWMENT FOR THE ARTS

NATIONAL SCIENCE ADVISORS PROGRAM OF THE OFFICE OF NAVAL RESEARCH
DISCUSSION

BIBLIOGRAPHY

APPENDIX A: October 12, 1995 Conference

EXAMINING FEDERAL APPROACHES OUTSIDE THE DEPARTMENT OF EDUCATION TO IDENTIFY AND DISSEMINATE THE BEST

Karen Bogart, Ph.D.
EduTech Limited

May 1997

INTRODUCTION

On October 12, 1995 the Federal Office of Educational Research and Improvement (OERI) sponsored a conference to examine what federal agencies outside the Department of Education were doing to identify and disseminate what was best in their respective fields (See Appendix A for transcript of this meeting). The impetus for this conference, in the words of Dr. Laurence Peters, formerly Counsel to the Select Education Subcommittee that authored the OERI reauthorization legislation, was the evidence that "Federal Education Research and Development is under appreciated in terms of budget and in terms of the way the public sees its operation." The wealth of material potentially available through Federal R&D is inaccessible if only because there is so much of it — curricula, research reports, analyses of educational policy and practice — but often little clue as to what works and under what conditions, and few guidelines for assessing what is potentially replicable by different schools and communities.

OERI's Educational Resources and Information Center (ERIC), for example, provides access to an enormous quantity of published material, including curricula and other innovations, but it is difficult for consumers to ascertain what is likely to have a positive impact based on their needs. Peters noted that the National Diffusion Network (NDN) has approved exemplary curricula and innovations in schools across the country, making funds available for schools interested in replicating that particular innovation. It has been an excellent system and it has made a great impact on schools throughout the United States. However, there is a problem. "Many of the programs that were exemplary or could be so classified could not get through this enormous pipeline to be approved by a distinguished board of people (Peters, 1995)." The October 1995 Conference represented an effort to learn from the examples provided by other federal agencies how effective programs, policies, practices and products could be given some kind of status that would make sure more people around the nation could take advantage of these resources.

The review presented in this paper continues that effort. This review first examines approaches used by the Department of Health and Human Services' Office of Adolescent Pregnancy Programs (OAPP) to identify what works in the field of adolescent pregnancy, parenting and prevention. OAPP relies, above all, on summative or outcome evaluation that is quantitative in nature and that uses a core set of common health indices, especially (1) onset of first sexual activity, (2) birth weight and (3) repeat pregnancy, to evaluate the effectiveness of prevention, pregnancy and parenting programs it funds. Particular attention is directed to several resource compendia to best practices, developed and/or disseminated under the auspices of DHHS.

These resources are available through several outstanding clearinghouses, especially the National Center on Education in Maternal and Child Health (NCEMCH), located at Georgetown University. This review also addresses a number of issues disclosed in discussions with nationally know experts in adolescent childbearing that are not directly addressed in existing guides to best practices, although these are critical issues that must be addressed if welfare reform is to achieve its goals.

Second, the review focuses on the activities of the Federal Emergency Management Agency (FEMA). FEMA's Preparedness, Training and Exercise Directorate - State and Local Preparedness Division (PT-SL) is engaged in "a massive experiment in technology" involving Internet and Distance learning through which it is harnessing the newest information technology in the service of an Information Clearinghouse activity. The Information Clearinghouse makes it possible for FEMA both to identify what works in emergency management preparedness and to disseminate best practices to its customers (states, regions and the international community, including thousands of emergency management workers) both electronically and in hard copy. FEMA's activities appear to be especially relevant to the Office of Educational Research and Improvement because they are specifically focused on identifying what works and disseminating the best.

Third, brief reviews are included of the activities of several other agencies to identify what is best in their respective disciplines. First among these other agencies are the Substance Abuse and Mental Health Services Administration (SAMHSA) and the Agency for Health Care Policy and Research (AHCPR), two other agencies of the Department of Health and Human Services. The formidable tasks before these federal agencies are ones of sifting through numerous research studies and reports (1) to identify effective health policies, treatments, intervention strategies and (2) to develop clinical guidelines that facilitate clinical practice in the health professions (AHCPR) and prevent and treat substance abuse (SAMHSA). While in the past these agencies have utilized consensus panels of experts and lay persons with great effectiveness, they have more recently shifted focus to using panels to assist in the development of "rules of evidence" that can be used to assess different methodologies and outcomes. Both consensus panels, extensively and ably described in a previous OERI report by Lois-ellin Datta (1994) and "rules of evidence" may offer suggestions to OERI that can be incorporated into the Expert Panel Conceptual Framework (Findbest System) that it is currently developing for finding the best in Education. A detailed review of the activities of NIDA are also provided in another review (Barkdol, 1996).

Finally, the activities of the Office of Naval Research and the General Accounting Office are briefly examined. The Office of Naval Research sends Science Advisors into the field and onto ships to assist Navy Officers to adapt and assess mature technologies developed by the civilian and military workforce of the Office of Naval Research. By analogy, OERI might consider sending educational advisors (or contractors) to assist local school districts most in need of assistance to adapt educational innovations, especially those involving advanced technology, such as Internet applications. The General Accounting Office periodically evaluates what works and does not work in different program areas and also has published a variety of documents intended to enable users to conduct their own quantitative and qualitative evaluations.

METHOD

Although different federal agencies seek to provide cost-effective information about programs, policies and practices that are best, they vary in the assessment approaches they use to identify the best and to make this information available to others. This in turn reflects their somewhat different interpretation of the ways in which they can best serve the public welfare. An overview of each agency's approach will be provided using a template that includes several dimensions along which the agencies differ:

Expert Panels. Federal agencies differ in the extent to which they rely on expert panels as well as in the ways in which they use panels. At one end of the continuum are agencies that rely on expert panels to review research evidence and formulate policy, especially in the field of health. Expert panels may also be used to establish rules of evidence, specific criteria that programs and practices must meet. Expert panels are also used to screen for success of a policy or program in at least one site. At the other end of the continuum are agencies that do not use panels at all, preferring instead an empirical approach in which the evidence is examined directly without mediation by an expert panel.

Quantitative and Qualitative Performance Indicators. Agencies differ in the extent to which they rely on quantitative performance indicators vs. qualitative performance indicators. Some agencies use quantitative indicators only while others have come to recognize that best results are achieved by using a combination of quantitative and qualitative indices to measure effectiveness of programs and practices.

Research Evidence. Federal agencies differ in the extent to which they have formal approaches for sifting through research evidence. Agencies within the Department of Health and Human Services not only establish criteria or rules for evaluating research evidence but also fund centers that will review the evidence and prepare reports to both professional and lay public. Other agencies have less concern about the possibility that a practice or program they recommend will be found wanting.

Promising and Exemplary Programs. Although all agencies evaluate programs and practices, some agencies provide guidance based on the results of the evaluation while for others consumers must weigh the evidence and decide for themselves whether to adopt a policy, program or practice.

Resource Materials. Federal agencies differ in the ways in which they disseminate information to the general public. Some agencies publish resource directories or other materials. The directory may include all the programs an agency has funded or only those programs that have met specified evaluative criteria regardless of funding source.

Information Technology. Federal agencies differ in terms of their reliance on information technology to identify and disseminate the best. At the high technology end of the continuum is the Federal Emergency Management Agency, which uses the latest information technology to call for nominations of outstanding emergency management practices and to disseminate information about emergency management practices nationally and internationally. Other federal agencies also make use of information technology by placing the full text of documents of all kinds on the Internet and providing user-friendly search engines. However, it is not clear how widespread the use of Internet is by consumers to obtain information about best practices. These efforts may be better viewed as directed at future widespread use of information technology rather than current. At present, even in universities, use of information technology to search for information about innovative practices seems to be fairly limited.

A caveat is in order before turning to a review of agency approaches to identifying and disseminating the best. There is always the possibility of exaggeration in the claims that a program, policy or practice may make about its success. However, scarce resources make it difficult if not impossible to conduct site visits to most programs to check out their claims about the services they offer, the numbers they serve, or the results they achieve. None of the federal agencies reviewed addressed this problem. There is a need for site visits or some other form of independent verification of the programs that resource materials are recommending. If consumers discover on their own that some program descriptions are distortions, they may not trust any of them. A review of federal agencies follows.

REVIEW OF FEDERAL AGENCY APPROACHES

OFFICE OF ADOLESCENT PREGNANCY PROGRAMS: DEPARTMENT OF HEALTH AND HUMAN SERVICES

Federal Agency	Office of Adolescent Pregnancy Programs
Target population	Pregnant and parenting adolescents, health professionals, social workers, educators
Uses online information technology to identify best	No
Uses online information technology to disseminate best	Indirectly. DHHS funds National Clearinghouse on Education in Maternal and Child Health at Georgetown University, which conducts online searches for users and provides extensive information to resource compendia, people, etc.
Uses consensus/expert/screening panels	No
Uses rules of evidence	No
Evaluation approach	Quantitative measurement, with a focus on health indices, including age at onset of first sexual activity (prevention programs), birth weight, incidence of repeat pregnancy (programs for pregnant and parenting adolescents).
Publishes hard copy resource materials	Yes.

KEY OBSERVATIONS:

- Best practices are measured by performance on a core set of health indices, especially, birth weight, incidence of repeat pregnancy and onset of first sexual activity.

- Quantitative methods such as multivariate analysis, are used to assess the effects of adolescent pregnancy programs on the core set of health variables.

- A qualitative study, "Adolescent Voices", is now in progress, sponsored by the Office of Adolescent Pregnancy Programs (OAPP).

- There appears to be consensus among experts as to which are the best school-based and school-linked adolescent pregnancy programs in the nation. Best programs appear to perform reasonably well on the health indices identified above but there are other additional criteria that the programs appear to have in common. These include (1) a holistic approach that includes prenatal care, on-site childcare, parent education, and case management services, and (2) a dynamic and charismatic program leader who wins the confidence and trust of adolescent parents.

- Adolescent pregnancy program directors acknowledge the need to measure other outcomes, in addition to health indices, including those associated with education and employment of the adolescent parents and those related to early childhood development of the infants, toddlers and preschoolers born to adolescent parents. However, educational and economic indicators are not the focus of OAPP.

- Resource directories to adolescent pregnancy programs do not, for the most part, report evaluation outcomes although they do indicate whether an evaluation is ongoing or completed. The resource directories also do not differentiate promising and exemplary programs. Consumers must judge for themselves whether a program is worthy of replication.

IMPETUS FOR ADOLESCENT PREGNANCY PROGRAMS

Every day in America:

2,795 teenagers get pregnant
1,295 teenagers give birth
7,742 teenagers become sexually active

Nationwide:

I. One million adolescent females become pregnant every year.
II. Approximately 519,000 teenagers give birth every year.
III. Nearly 1 in 5 teen mothers become pregnant again within one year.
IV. Approximately 60% of teen mothers have histories of sexual and physical abuse, primarily by adult men averaging age 27.
V. Among mothers ages 11-12, fathers average 9.8 years older, among ages 13-14, 4.6 years, and among ages 15-19, 3.7 years.

(From "A STEP-UP for Arvada's Teens", Jefferson County Department of Health and Environment, Arvada, CO; data on number of teen mothers giving birth every year from *The Washington Post Parade*, January 1997).

For reasons such as these, attention needs to be paid to the triple needs of (1) providing pregnant adolescents with prenatal care, (2) teaching adolescent parents how to stimulate the cognitive, emotional and social development of their infants, toddlers and preschoolers and (3) providing teen parents with the education and occupational skills that will enable them to avoid or escape the poverty, substance abuse, hopelessness, learning disabilities, successive generations of teen pregnancy, and criminal activity so often associated with adolescent childbearing. In response to these needs, the Office of Adolescent Pregnancy Programs and the Maternal and Child Health Bureau have both funded adolescent pregnancy programs with evaluation components, and supported the development and dissemination of resource materials that provide information about best practices.

Approximately 131 model school-based and school-linked adolescent pregnancy programs are described in several different resource compendia to best practices including:

- *Adolescent Pregnancy Prevention: A Compendium of Programs* published by the Healthy Mothers, Healthy Babies Coalition (1995), associated with the American Association of Obstetricians and Gynecologists;
- *The Adolescent Family Life Demonstration Projects: Program and Evaluation Summaries* published by the Office of Adolescent Pregnancy Programs (1990);
- *School-Based Programs for Adolescent Parents and their Young Children: Overcoming Barriers and Challenges to Implementing Comprehensive School-Based Services* published by the Center for Assessment and Policy Development (CAPD) (1994);
- *School-Based Programs for Adolescent Parents and Their Young Children: Guidelines for Quality and Best Practice* published by the Center for Assessment and Policy Development (1996).

The programs outlined in an additional compendium were also reviewed for this report, although these programs are not included in the discussion of 131 school-based and school-linked programs that follows since these are currently funded and no evaluative data is as yet available on them.

- *Adolescent Health Report* published by the Maternal and Child Health Bureau, Health Resources and Services Administration (1996).

In addition, a snowball approach has been used to conduct interviews with key individuals in an effort to ascertain the extent to which the available resource materials describe best practices that accurately and completely address health, educational and social needs of pregnant and parenting adolescents. Individuals interviewed include:

Dr. Patrick Sheeran, Director, Office of Adolescent Pregnancy Programs (OAPP);
Ms. Elizabeth McDonald, Evaluator, OAPP ;
Dr. Dennis McBride (consultant to OAPP and quantitative evaluator specializing in health outcomes);
Dr. Teresa Okwumbua (Member of the Board of the National Organization of Adolescent Pregnancy, Parenting and Prevention; Adolescent Pregnancy Program expert in Memphis City Schools, TN) ;
Ms. Donna Butts (Executive Director, National Organization of Adolescent Pregnancy, Parenting and Prevention or NOAPPP);
Dr. Sharon Enright, Director, GRADS program, Ohio Department of Education - a school-based parent education curriculum, originally launched in Ohio and now disseminated nationally;
Ms. Susan Batten, Center for Assessment and Policy Development (CAPD) based in Bala Cynwood near Philadelphia, which provides resource materials on best practices and technical assistance to school-based programs seeking to develop adolescent pregnancy programs;
Dr. Sandy Dixon, Principal, New Futures School in Albuquerque, NM - highly recommended alternative public school, also author of printed materials which she is sending;

Dr. Rosetta Stith, Paquin School, Baltimore, MD.
Ms. Amy Loomis, former Director, San Francisco Teen Pregnancy, Parenting and Prevention Programs, currently President, Marin Creative Options;
Ms. Rhonda Simpson, State Coordinator of Adolescent Pregnancy Programs, California State Department of Education;
Ms. Charlene Clemans, Program Director, Teenage Adolescent Pregnancy Programs, San Francisco, California;
Ms. Karen Holecker, National Coordinator, Parents As Teachers;
Ms. Sarah Fletcher, Director, Teen Pregnancy Programs, Elkins, West Virginia.

METHODOLOGICAL CONSIDERATIONS

Several methodological considerations must be kept in mind as information about best practices in adolescent pregnancy and parenting is reviewed:

By program is meant both the activities of what may be a comprehensive or alternative school (actually encompassing many different programs) and specific approaches, such as a nationally disseminated curriculum, that is in fact offered as a component in many different schools.

There may be other outstanding programs – perhaps a few, possibly many. They may not have come to the attention of experts in the field or developers of resource compendia because they are local efforts, including those initiated by schools and school districts, that have not sought federal or foundation funds or otherwise networked or formed linkages with national organizations. Practitioners are often too busy running programs to write about them.

The majority of approaches described by the *Healthy Mothers/Healthy Babies Resource Compendium* and the *Office of Adolescent Pregnancy Programs Resource Compendium* are lodged in community-based organizations, including social services agencies, health centers, non-profits and local branches of national organizations such as Planned Parenthood, which often conduct programs at least in part in collaboration with comprehensive or alternative public schools. Many, if not most, are privately owned and operated. Some are County or local Departments of Health and, more rarely, school districts. Whether this is true in general of adolescent pregnancy programs or only characteristic of those described in these compendia, with their focus on maternal and child health, cannot be determined at this time. Major components of these programs are provision of health care (OAPP programs) and health education and HIV/AIDS education (Healthy Mothers/Healthy Babies programs).

The Center for Assessment and Policy Development (CAPD)-Resource Compendium, in contrast, identifies best practices based in @20 schools and school districts nationwide. In a conversation with Susan Batten of CAPD, she reports that an initial search by CAPD showed so little evidence of best practices in adolescent pregnancy programs that CAPD decided to shift its own initial effort from evaluation of best practices to provision of technical assistance to introduce what CAPD believes are best practices, based on its review of existing programs.

CHARACTERISTICS OF SCHOOL-BASED AND SCHOOL-LINKED ADOLESCENT PREGNANCY PROGRAM CITED AS BEST PRACTICES

Who Is Served by Adolescent Pregnancy Programs. Adolescent pregnancy programs are distributed across the nation and serve inner city, urban, suburban and rural communities. Most programs serve both adolescent mothers and fathers. About one third provide prenatal care. Approximately one third also provide child care services for infants and toddlers, occasionally with an early childhood education program. Somewhat fewer serve preschoolers. Where programs are lodged in alternative schools, single sex instruction is common with classes restricted to teen mothers; teen fathers participate in after school programs. Contrary to popular belief, the population served by the largest number of adolescent pregnancy programs is White non-Hispanic followed by African Americans, and to a somewhat lesser extent, Hispanics. Fewer programs serve Asians and Native Americans. A few programs address the needs of the state as a whole and a few are national in focus: many of these operate in multiple sites, for example, sending parent or health education teachers and case workers into different schools and communities to teach particular curricula. The Grads Program, originating in Ohio, and Parents As Teachers, originating in Missouri, are examples of curriculum-based programs classified as best practices by the National Dissemination Network that are taught across the country.

Single vs. Multiple Site Delivery of Services. Some adolescent pregnancy programs deliver their services in a single site. Single site programs are typically located in comprehensive public schools (especially high schools), alternative public or private schools, and non-residential community facilities, such as a social service agency or a community center. Some single site programs are also located in residential facilities. More than half the programs described in these resource compendia, however, involve multiple forms of intervention and are delivered in multiple settings, with adolescent parents traveling between them. Teen parents are especially likely to travel for health care.

Public vs. Private Providers of Services. Most providers of Adolescent Pregnancy Programs are private non-profits, including Parents As Teachers, the YWCA and Planned Parenthood, all of which offer adolescent pregnancy programs in many different states. Even when they are private non-profits, however, they commonly use community facilities and public schools to deliver their services, at least in part. As a result, they are typically working in collaboration with schools and school districts.

Prevention Only Programs vs. Care and Prevention. Out of 131 programs identified in resource compendia reviewed, @50 provide prevention education only and are aimed at adolescents who have never been pregnant. The overwhelming majority of the other 80 offer both care and prevention services. Research suggests that those girls most at risk of pregnancy, moreover, and therefore most in need of pregnancy prevention programs, are the siblings of teen mothers. However, prevention is a critical issue not only for girls who have never had a baby but also for adolescent mothers who have already delivered a child. Research suggests that it is the second pregnancy in adolescence, rather than the first, which overwhelmingly reduces opportunity to escape poverty and other negative consequences of adolescent childbearing.

Services Offered. The services identified by practitioners as most critical to best practice include on-site childcare in school-based programs, which is all the more important because many teen mothers hesitate to leave their babies with strangers; prenatal care and postnatal health care; case management services; and home visiting. School-based programs also provide academic and vocational education, as well as parenting classes and health education. Case management and home visiting are particularly important for the 50% of teen mothers who drop out of school and who have not shown as yet an inclination to return, the new requirements of welfare reform that adolescent mothers be in school to receive benefits not withstanding. California has pioneered in making case management and other services available to pregnant minors whether or not they are in school (cf. Simpson, Clemans).

There is consensus on the part of experts that the best programs provide on-site child care, early childhood education, and health care for the family. There appears to be agreement that the best programs are holistic alternative schools, addressing a teen mother's educational needs by addressing all her needs and those of her baby as well as her extended family including siblings and grandparents. However, these observations are based on interviews with experts in the field, rather than on the available resource directories, which provide limited guidance as to the best of best practices. In practice, there are few alternative schools providing multiple services across the nation. Interestingly, in addition, the fact that a program is delivered, at least in part, in a school does not guarantee an academic component. Some programs delivered in schools provide parent education or health education but do not have an academic component that provides or enhances educational instruction. Other programs that do not use the school as a setting may nonetheless have strong academic or vocational educational components, for example, offering job training, academic tutoring, or career counseling. Further, many of the programs described in the newest of the resource compendia — Healthy Mothers, Healthy Babies — include such components as peer support groups and mentoring, while none of those described in the oldest of the resource directories — the Office of Adolescent Pregnancy Programs Resource Compendium — are described in terms of these components.

Evaluation. The overwhelming majority of programs report that evaluations of their programs are in progress or had been completed. Whether the evaluation is quantitative or qualitative, formative, short term outcome or long term impact, is sometimes reported or can be inferred. Commonly, adolescent pregnancy programs are evaluated in terms of a small number of quantitative criteria, especially age at onset of first sexual activity (prevention only programs), incidence of low birth weight, and incidence of repeat pregnancy. However, the results of the evaluation are almost never reported in any of the resource directories reviewed. Moreover, even when results are reported, no inference is possible as to whether this program is promising or exemplary, however these terms may be described. Insofar as indices are identified in the resource compendia (and for the most part, they are not), these are likely to be the numbers of adolescent mothers, fathers and infants served. The evaluation data described would not be likely to be of much use to a potential consumer interested, for example, in replicating one of these programs. However, it is possible that program directors would provide more information on request.

ADOLESCENT PREGNANCY AND PARENTING NEEDS NOT ADDRESSED IN BEST PRACTICES RESOURCE COMPENDIA

A variety of issues and needs identified in literature reviews and in discussions with experts in the field are not reflected directly in the quantitative indicators used to assess best practices in adolescent pregnancy programs. In fairness to OAPP, these other indicators do not directly reflect its mission. Discussion with school principals and program directors, however, reveals that they are fully aware of these needs and do address them in their programs, although not always in the same way. Since these needs might affect a consumer's choice of program to replicate, they should not be ignored. Issues include the following:

Demographic Trends. These include basic demographic changes such as:

- **Incidence of Adolescent Pregnancy.** The incidence of teen pregnancy has declined somewhat in the past year (cf., Child Trends, 1996); however, the ratio of births to unmarried teen mothers has continued to increase dramatically; it was 78% nationally in 1994, the most recent year for which there are figures, and @98% in the District of Columbia (cf., Daniel Patrick Moynihan, Washington Post, January 1997). A few programs identified to date, however, such as the New Futures School, an alternative public school in Albuquerque, NM, address 80% of the need in the state. However, even when outstanding school-based and school-linked adolescent pregnancy programs are available in a community, they may only address a fraction of the need.

- **Race/ethnicity of Pregnant and Parenting Adolescents.** Adolescent childbearing is a problem that increasingly affects every ethnic group. Among some racial/ethnic populations, such as the Hmong, cultural values strongly support childbearing in adolescence.

- **Mother's Age.** Pregnant and parenting adolescents in some communities are younger than ever before. The Margaret Hudson Program, in Tulsa, OK noted that whereas a few years ago it was serving students 16 and older, it is now mostly serving adolescents 15 and younger, with the youngest girls becoming pregnant at 9 or 10 and delivering at 10 or 11. Further, hospital records about the age of the adolescent mother are sometimes incorrect, since the mother is not required to provide a copy of her own birth certificate when delivering. She may report that she is several years older than she actually is and her actual age may go undetected, especially if her physical development is similar to that of an older adolescent.

- **Father's Age.** Increasingly, fathers are out of their teens. Two thirds of the fathers are reported by different experts to be in their twenties.

- **Grandparents' Age.** There is a possibility, especially with younger adolescent mothers, the fathers may be boyfriends of mothers who may be only 28 or 32 years old themselves. In some cases, the parent of the adolescent may herself be in a stage of arrested adolescent development in which she views her own daughter as a competitor for the affections of the boyfriend. In other cases, the mother may depend on the boyfriend to provide rent and food. Both kinds of situations make it difficult to intervene on behalf of the parenting adolescent, since the boyfriend is likely to remain in the home.

- **Siblings of Adolescent Mothers.** Siblings of adolescent mothers are reported to be at greatest risk of also having a child while in their teens.

Critical Needs Not Addressed by Indicators Used to Assess Best Practices. Other issues identified by experts in the field call attention to the need to address issues such as the following in assessing in assessing best practices. Experts report the following:

- **Multiple populations of adolescent parents:** Experts (in interviews) agree that there are multiple populations of pregnant and parenting adolescents with different needs. These include:
  - Older first time adolescent mothers (16 or 17 and older) who are fairly mature, for whom adolescent motherhood may actually be an adaptive response to an impoverished environment that does not offer many opportunities for gratification other than motherhood, and who are not likely to have a repeat pregnancy while in their teens;
  - Younger first time adolescent mothers (14 and younger) who are more likely to have been victims of child and sexual abuse in early childhood, for whom pregnancy is a maladaptive response, who are often bitter, rebellious and resistant to efforts to help them, who are more likely to have a low birth weight baby and who are most likely to have one or more repeat pregnancies while in adolescence;
  - Substance abusing students;
  - School drop outs returning to school, especially as a result of recent Welfare Reform;
  - Different racial/ethnic populations with different cultural norms about the desirability of childbearing in adolescence; and
  - Middle class adolescents who will complete high school and enter postsecondary education.

- **Educational needs of adolescent parents.** Adolescent mothers have significant educational needs. At one end of the spectrum are those who are educationally behind their age cohort in school. The majority of adolescent parents comprise this population. Many did not bond with school in first or second grade. They started behind, fell further behind, and never caught up. They may be reading at a second grade level. These factors do not bode well for their occupational prospects, particularly given the increasing reliance of all levels of the workforce on advanced technology. Even industrialized sewing – so often thought of as "women's work" – today demands computer skills. Lacking in computer skills and in the reading proficiency necessary to master computer skills, teen mothers are often relegated to the lowest paying jobs. Such jobs mean that adolescent mothers cannot even provide for themselves and their babies the minimum standard of living formerly available to them through welfare.

  More broadly, educational and occupational prospects for pregnant and parenting adolescents are perceived by the small number of experts interviewed to date to be limited, above all, by the inadequate access that these students, so many of whom are low income, have to the same kind of technological expertise that middle class students are developing — through access to the computer, multimedia and the Internet. Unless they have access to computer education and, more broadly, to technology as a tool, in a society increasingly dependent on science and rapidly changing technology, their economic opportunities will continue to be marginal. The need for technological education, both as a means to an end (using computer technology, including the technology of Distance Learning and of self-paced instruction, to correct educational deficits) and as an end in itself (improving economic prospects by learning computer skills), in adolescent parenting programs, is overwhelming.

Adolescent fathers may have more options than mothers in terms of employability, because there are more reasonably well paying "traditionally male" jobs that may not require computer skill, such as truck driver, garbage collector, plumber or landscaper. At the other end of the intellectual continuum, moreover, are adolescent parents who are gifted and talented, but who are left out of the programming for gifted and talented populations.

Low income adolescent mothers, further, are often counseled into leaving school and obtaining a GED instead of a high school diploma. Not only may they be unlikely to complete the GED but this degree effectively closes the door to continuing education. Middle class adolescent mothers are more likely to obtain postsecondary education whether they obtain a high school diploma or GED.

- **Zero tolerance.** Many of the school-based adolescent pregnancy programs identified to date address the needs of "healthier" adolescent parents, since they usually serve teen parents who have opted to be there and often only a fraction of those, due to limited resources. They may accept substance abusing adolescent mothers and make a concerted effort to change their behavior. They may also require students to be in attendance or forego class credit. At the same time, they practice zero tolerance for such behavior as fighting and carrying a weapon. In addition, the neediest or most rebellious adolescents may either not learn of these programs, choose not to participate in them, or be dropped from them when they prove hard to handle. Half the teen mothers in the nation are continuing to drop out of school, welfare reform notwithstanding. They drop out and refuse to return, above all, because they are far behind their age cohorts in educational level. However, adolescent parents who drop out must not be ignored, if only to protect the infants, toddlers and preschoolers who may remain in their custody. The state of California, which has the highest incidence of adolescent pregnancy in the nation, is noteworthy because it does not practice zero tolerance. Pregnant minors receive case management services whether they are in school or not..

- **Sexual and physical abuse of adolescent mothers.** Experts who serve adolescent mothers, especially school principals, report (in interviews) that one of the most critical issues in need of redress is sexual and child abuse, especially of girls. These children are at greatest risk for early pregnancy and, in most cases, when a 9 or 10 or 11 year old is pregnant, it is because of sexual abuse and rape. School principals report considerable unwillingness (including Child Protective Agencies) to address these issues because it is easier to blame the girl herself for her pregnancy than to take community action to protect these children.

- **Early childhood education needs of children born to adolescent mothers.** Recent research on the early development of the brain in infancy and early childhood (cf., White House Conference on Early Childhood Development, April 1997) calls attention to the overwhelming importance of positive stimulation for the developing brain in the earliest weeks, months and years of life. For this reason alone, there is a critical need for early childhood education programs for the infants, toddlers and preschoolers born to adolescent parents. Whether due to lower birth weight, poor nutrition, or substance abuse during pregnancy, the children born to adolescent mothers have poorer developmental outcomes than children born to older mothers. One outstanding program, Parents as Teachers, serving 25,000 adolescent mothers in 1900 sites around the country, teaches parent educators to go into the home and work with parents on a one to one basis to stimulate cognitive development in their infants and toddlers. All too few programs, however, even those serving pregnant and parenting adolescents, provide the early childhood education that may be especially critical for children born to adolescent mothers. Some successfully address the issue of low birth weight through prenatal health care and health education, but provision of service seems restricted to physical health. In connection with these and other services, experts noted that they were often combating hostile public sentiment, based on the perception that any form of "help" would encourage further childbearing.

  Related to issues of early childhood education are those of foster care and adoption. All too many children raised in foster care become adolescent mothers producing babies who will similarly be raised in foster care. Clinton has called for an increase in adoption of children now raised in foster care. Children born to adolescent parents may be at greater risk of developmental delays and other negative consequences whether they are raised by their birth parents, in foster care, or by adoptive parents. Early childhood education programs tailored to the needs of children born to adolescent parents (and others at risk of development delay) may be especially critical whether these children are raised by their birth parents, in foster care, or by adoptive parents.

- **Need for further evaluation of adolescent pregnancy programs.** Some evaluative data exists for school-based and school-linked programs addressing adolescent pregnancy, parenting and promotion. The data are, for the most part, quantitative rather than qualitative and are summative studies that focus on health indices measuring short-term outcomes rather than long term impact. These data demonstrate that

  (1) school-based and school-linked programs providing prevention education do delay the onset of sexual activity;
  (2) adolescent mothers receiving prenatal care deliver higher birth weight babies than those who do not receive prenatal care;
  (3) younger adolescent mothers are at more risk of delivering a low birth weight baby than older adolescent mothers;
  (4) older adolescent mothers participating in adolescent pregnancy and parenting programs are less likely to have a repeat pregnancy than younger adolescent mothers in these programs.

While there are studies that evaluate school-based and school-linked adolescent pregnancy programs in terms of health indices, there appear to be few studies that assess the effects of an adolescent pregnancy program on educational and occupational outcomes for adolescent mothers (and fathers). In part, this may be an artifact of funding; insofar as adolescent pregnancy programs have had federal funds to date, these have been awarded by the Department of Health and Human Services, which may explain the health focus of evaluation data and, in fact, of the services provided by the programs themselves. Principals of school-based adolescent pregnancy programs interviewed to date also report that they have not had the financial resources to conduct studies of educational and occupational outcomes, especially insofar as these would require long-term follow-up of students after they leave the program. There are also few evaluations of the effects of the early childhood education programs that some adolescent pregnancy programs provide for the infants, toddlers and preschoolers, although it would seem likely these programs would make an important difference in the development of children born to adolescent mothers. In the case of some outstanding school-based programs, principals report that they are tracking graduates of their programs longitudinally but lack the funds to analyze the data they have been collecting.

Although there appears to be an absence of studies of the effects of school-based and school-linked adolescent pregnancy programs on educational and occupational outcomes as well as on early childhood development, there is no lack of studies, journal articles, scholarly books and media attention testifying to the high cost of adolescent childbearing itself — on adolescent mothers, on adolescent fathers, on the children born to adolescents, and on society as a whole especially in inner cities and urban communities. These studies call attention to the overwhelming costs associated with adolescent pregnancy and parenting, in terms of reduced educational and occupational opportunity especially for adolescent mothers (and particularly after the birth of a second or third child more than after the birth of a first child) and in terms of the higher incidence of learning disabilities, development difficulties, substance abuse and criminal behavior among children raised in the (financially, educationally and socially) impoverished environments that often characterize adolescent parenting.

EXAMPLES OF OUTSTANDING ADOLESCENT PREGNANCY PROGRAMS

The resource compendia reviewed provide descriptive information on more than 131 programs. However, they do not always make clear at the outset the criteria that have been used to select programs for inclusion. Nor do they provide any information that would permit the differentiation of promising from exemplary programs. Some school-based programs that have received widespread acclaim are not included in any compendium; they have been identified through contacts with key persons in the field.

Examples of programs that have received widespread acclaim include:

- The Laurence G. Paquin Junior-Senior High School for Expectant Teenage Mothers in Baltimore, Maryland: alternative public school serving 800 adolescent mothers annually and including on-site childcare and early childhood education programs for infants, toddlers and preschoolers who can remain in the program until 5 years of age. Dr. Rosetta Stith is Director. Support is provided by Baltimore City School District and supplemented by federal and foundation grants. This program is not abstracted in any compendium.

- The New Futures School in Albuquerque, New Mexico: alternative public school serving 500 adolescent mothers annually and including on-site childcare and early childhood education programs for infants and toddlers to 3 years. Dr. Sandy Dixon is Principal. Support is provided by the State of New Mexico, which returns tax dollars to local school districts for school programs. This program addresses 80% of the need in the State of New Mexico. The other 20% are school drop-outs. This program is abstracted in the CAPD Resource Compendium published in 1994..
- The Margaret Hudson School in Tulsa, Oklahoma: 3 privately operated (non-profit) alternative schools serving 400 adolescent mothers and school-linked services in comprehensive high schools for an additional 150 adolescent mothers annually: this program addresses only a fraction of need in Tulsa, where as many as 1000 adolescent mothers are reported each year. Ms. Jan Figurt directs this program, which is not abstracted in any of the compendia.
- School-based and school-linked programs in the state of California: California has been a leader in providing school-based and school-linked services to pregnant and parenting adolescents, as well as services to the 50% of teenage mothers who drop out of school. Although California has led efforts to serve pregnant and parenting adolescents as well as to promote prevention, more than 100,000 adolescents give birth every year. Its efforts notwithstanding, in California in particular, there are some recently immigrated subgroups of the population that encourage adolescent childbearing, such as the Hmong and Hispanic populations. In addition, California's estimates are inflated by inclusion of 18 and 19 year old women as adolescents. Ms. Amy Loomis, Ms. Rhonda Simpson and Ms. Charlene Clemans are nationally known experts who have pioneered in the development of adolescent pregnancy programs in California.

AVAILABILITY OF CONSUMER-ORIENTED INFORMATION IN ADOLESCENT PREGNANCY, PARENTING AND PREVENTION

The resource compendia that provided the basis for part of this review were identified through the Department of Health and Human Services search engine supplemented by guidance from the federally funded National Clearinghouse on Education in Maternal and Child Health. The information is not as readily available, however, as it could be to potential consumers, including educators, health professionals, social workers, parents, and pregnant and parenting adolescents. Diligent searching is required to identify the resource compendia that have recently been published. Information of this kind could be made more readily available through Internet itself. Using Internet to identify and disseminate information about best practices is a primary focus of the next agency to be reviewed, the Federal Emergency Management Agency.

HARNESSING INFORMATION TECHNOLOGY TO STREAMLINE GOVERNMENT: THE FEDERAL EMERGENCY MANAGEMENT AGENCY (FEMA)

Federal Agency	Federal Emergency Management Agency
Target population	States and local governments, voluntary organizations, business and industry
Uses online information technology to identify best	Yes
Uses online information technology to disseminate best	Yes
Uses consensus/expert/screening panels	Screening panels
Uses rules of evidence	No
Evaluation approach	Approach must have worked in at least one setting, evidence provided by community submitting practice for FEMA consideration
Publishes hard copy resource materials	Yes. First volume of Compendium of Exemplary Practices in Emergency Management published in 1996; second volume available in draft form.

KEY OBSERVATIONS

- The Federal Emergency Management Agency (FEMA) does not differentiate between promising and exemplary programs. The programs that FEMA calls exemplary would probably be considered promising using OERI criteria.

- FEMA relies on information technology as a key mechanism for disseminating information about exemplary practices in emergency management preparedness and for soliciting nominations of candidate programs and practices from the global emergency management community.

- Identifying and disseminating information about best practices to states and local communities is a key mission of FEMA's Information Clearinghouse.

- FEMA's screening committee reviews evidence of success in one site as reported by states and local communities that propose best practices for consideration. However, programs and practices are not reviewed rigorously.

IMPETUS FOR FEMA ACTIVITIES

The activities and goals of FEMA's Preparedness, Training and Exercises Division-State and Local Initiatives (PT-SL) Directorate are reported in detail because FEMA offers an outstanding example of the way one federal agency's approach to harnessing advanced technology to identify and disseminate the best the nation and international community have to offer in emergency management practices, at the same time that it leverages scarce resources and streamlines government. As described below, FEMA's PT-SL Division goes beyond harnessing online information technology to identify and disseminate the best in emergency management practice, to using the latest technologies of Distance Learning to provide training to thousands of emergency workers nationwide. FEMA is preparing for a not too distant future in which all schools and colleges, homes and businesses have Internet access and in which using the technology will, hopefully, be second nature to young and old alike.

FEMA has undergone a transition from nuclear preparedness to an all hazards approach. At the core of its mission is the creation of an emergency management system built on a partnership of local, State and Federal governments, voluntary agencies, business and industry, and individual citizens. Most disasters can be handled at the local or State level, and in fact, most states and regions are most knowledgeable about their own emergency management needs. However, once the President has declared a major disaster, FEMA coordinates the Federal response and provides recovery assistance.

More broadly, FEMA's goal is to enter into partnership with states and local governments, voluntary organizations, business and industry, internationally to identify outstanding emergency management practices. Through these partnerships, FEMA is able to leverage existing resources to collect, store, retrieve, analyze, evaluate and disseminate information about what works globally to a community of users nationwide. FEMA PT-SL capitalizes on the newest information technology, making its resources available online as well as in hard copy, provides technical assistance, conducts training institutes and awards grants to reduce risks. FEMA has also used strategic planning to develop goals and management processes that enhance its ability to transfer knowledge from developers to users, as part of its function as a "Reinvention Laboratory." By analogy, OERI could use the FEMA approach to build on what has been found to work in schools and school districts across the country and, possibly, around the world.

FEMA ONLINE

FEMA is engaged in a massive experiment to harness the newest information technology and use it to carry out its mission, by serving as an Information Clearinghouse for outstanding emergency management practices. All information is available on-line through FEMA's World Wide Web address. FEMA's homepage is perhaps the best that this researcher has reviewed, within or outside the federal government. This is no mean achievement, since the federal government in general tends to have very good information access through Internet homepages. FEMA information is easily accessible through FEMA's Homepage with hypertext links that make it easily possible even for inexperienced users to navigate from one source of information to another. FEMA has a search engine that makes searching easy, whether a boolean or simple search. Information about all components of FEMA are available through the FEMA Homepage and all publications can be downloaded from Internet. One of the best and first of these, for identifying and disseminating what works, is FEMA's resource directory, the Compendium of Exemplary Practices. As illustrated below, however, the Compendium is one component of an overall approach intended to leverage scarce resources through a partnership with states and local communities that provides thousands of emergency workers nationwide with emergency management preparedness training and resources. The Resource Compendium can be downloaded from FEMA's Web Page using Acrobat Reader. Emergency workers across the country can also nominate new approaches for inclusion in a later directory on-line.

A COMPENDIUM OF EXEMPLARY EMERGENCY MANAGEMENT PREPAREDNESS PRACTICES

FEMA has set about a task of leveraging existing and shrinking resources in the field of emergency management. What has emerged is a partnership in preparedness initiative with federal regions. The initiative began with a letter from Kay Goss, Associate Director for the Preparedness Training and Exercises Directorate to all 10 FEMA regions to solicit promising or exemplary practices from states in those regions. The initiative has produced a first volume available both in hard copy and on WWW: "Preparedness. A Compendium of Exemplary Practices in Emergency Management". Practices are indexed as to subject area, state in which they were originally developed and tested, and contact persons. Each program description includes, in addition to this information, program description, evaluation information, annual budget, sources of funding, target population, setting, and project startup date. Regulations are not published in the Federal Register and OMB clearance is not required.

FEMA's partners in the development of the Compendium have been State and Local Governments, educational institutions, the private sector and volunteer organizations, supported by annual Congressional funding provided through FEMA to the States. FEMA's job is to build a strong and effective emergency management infrastructure nationwide and through technology, such as Internet, globally. The initiative encompasses a linkage of national emergency management systems on a global basis.

The FEMA Compendium set out to describe public and private sector initiatives. The first edition includes approximately 78 practices around the country that have been screened and have worked in at least one setting. The Compendium refers individuals to the points of contact that have made the practices a successful reality. This publication, available on WWW as well as in hard copy, empowers the Emergency Management community to take existing talent, existing resources, and leverage them. FEMA is surfacing new ideas through its initiative that includes what is happening in the private sector, at the local level, in businesses, community groups. At the same time, it is important to note that, although there is a panel that screens approaches, they have not been subjected to rigorous evaluation. A practice is considered exemplary if one community can provide documented evidence of success in using it. Whether it has succeeded or failed in other settings is not relevant to this review.

Individuals and organizations are invited at the back of the hard copy Compendium and on Internet to submit emergency practices they consider exemplary. FEMA communicates with local practitioners and with individuals about each practice. However, FEMA itself does not evaluate practices. A screening panel whose members include representatives (peers) from various constituencies including State and local governments and the private sector assesses practices to assure that definition of exemplary practice is met: a practice that has worked in one place. Vernon Adler has said,"We are less concerned with how efficient it is, how effective, we leave that to individuals who use it. We are fostering partnerships and communication at all levels." In other words, the screening process is one of examining on an individual basis the empirical evidence for the success of a given practice. The process, of necessity, involves subjectivity but it is the subjective judgment of experts which provides the basis for inclusion or rejection. Stated in another way, the original development of a successful practice may have cost $250,000; the Compendium makes it possible for others to benefit from this practice gratis, through a process of technology transfer from the original development.

The Compendium is similar in its intent to other approaches intended to avoid duplication of existing practices and to promote adapting or adopting what has already been shown to work as an alternative to reinventing the wheel . What is especially noteworthy about this approach, however, is its dynamic nature: because the Compendium is available on Internet as well as hard copy, it can continuously be updated as users use and report back on the success of emergency management practices. More than 48,000 hard copies of this resource have been printed. Of this total, more than 46,000 have been sent out to various offices, directorates, regions, state governments, so that those in emergency management around the country have copies. A brief reader's survey form has been included; initial returns have been extremely favorable. A second edition is now being developed.

STRATEGIC PROCESS MANAGEMENT (SPM)

FEMA's effort to leverage existing resources is not limited to development of its *Compendium of Exemplary Practices*. The State and Local Preparedness Division within FEMA has been designated a "Reinvention Lab" within the Agency by the Director, James L. Witt. This Division was initiated in August of 1995, in response to the themes within the National Performance Review recommendation, to downsize the Federal government, increase responsiveness to customers, reduce costs and inefficiencies, and empower employees at all levels to simplify management structure . As a Reinvention Lab, the State and Local Preparedness Division operates within a team structure whose purpose is to leverage scarce financial and human resources available to FEMA to achieve transfer of knowledge.. This is an extension of the concept behind the Compendium of Exemplary Practices in Emergency Management.

The State and Local Preparedness Division in FEMA is using Strategic Process Management (SPM) concepts to improve Information Dissemination by analyzing work processes and uncovering complexity, duplication, and obsolescence. SPM has helped FEMA's Preparedness Division analyze how well the organization is currently designed to meet customer needs while finding opportunities for improvement and innovation. The staff assume greater responsibility for managing tasks by working in teams. As a result, strategic initiatives, goals, mission, processes, objectives, staff assignments and tasks have been reassessed. This reassessment is enabling the State and Local

Preparedness Division to meet its primary goal of improved information dissemination to more effectively serve State and local governments in the implementation of their emergency management duties.

SPM focuses on team structure. The team structure, designed as a part of the Strategic Management Process, leverages the resources available for improved response to FEMA's customers. One team is the Information Clearinghouse Team. The team provides technology policies and infrastructure.

GLOBAL EMERGENCY MANAGEMENT INFORMATION NETWORK INITIATIVE (GEMINI)

FEMA's information network is international. The ability of a community or a nation to respond to disasters and manage risk is critically dependent upon local, national and global information and resource networks. Unprecedented technology exists whereby all human emergency management knowledge could be instantly accessed to support all aspects of local to national emergency management. This would involve joint efforts of international organizations, government agencies, private industry, universities and interested citizens throughout the world. Each would contribute a share of information to create a virtual repository of emergency knowledge accessible to all through a global information infrastructure. More specifically, the objectives of GEMINI are to:

> Develop and implement systems to acquire, process, manage, display and disseminate information to support decision making for natural, technological, biological and humanitarian disaster responses, environmental monitoring and risk management.
>
> Develop and implement global networks to exchange information among emergency management organizations and the public in developed and developing nations.
>
> Establish global protocols for information standards, emergency access and search techniques to facilitate rapid exchanges of emergency management information.

TRAINING INSTITUTES

FEMA conducts training institutes for emergency workers, teaching emergency management skills. In addition to operating residential training institutes, FEMA has taken advantage of the latest technology to make training available cost-effectively to thousands of emergency workers through Distance Learning Centers. FEMA is planning to offer a diploma and university credit in emergency management.

BRIEF OVERVIEW OF APPROACHES TAKEN BY SELECTED OTHER FEDERAL AGENCIES

DEVELOPING RULES OF EVIDENCE: THE SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION

Federal Agency	Substance Abuse and Mental Health Services Administration
Target population	States and local governments, lay public
Uses online information technology to identify best	No
Uses online information technology to disseminate best	No
Uses consensus/expert/screening panels	Yes
Uses rules of evidence	Yes. Expert panels establish "rules of evidence" for each prevention approach investigated.
Evaluation approach	Research and practice are carefully reviewed for specific prevention approaches and then assessed for effectiveness.
Publishes hard copy resource materials	Yes. Three types of documents are published for each prevention topic. These include Guidelines, Practitioners Guides and Informational Brochures. The first guidelines are to be released shortly.

KEY OBSERVATIONS

- The Substance Abuse and Mental Health Services Administration (SAMHSA) does not differentiate between promising and exemplary programs.
- SAMHSA uses expert panels.
- SAMHSA develops rules of evidence.
- SAMHSA publishes different types of documents for different levels of consumers.

SAMHSA APPROACH TO IDENTIFYING WHAT WORKS

While FEMA is pioneering in the use of the latest in information technology to identify and disseminate the best in emergency management preparedness, SAMHSA is forging ahead in the development of "rules of evidence" that can be used to evaluate the best in substance abuse prevention and treatment. Its approaches may a model to OERI for establishing meaningful criteria against which individual programs, policies and practices can be evaluated, as well as for developing and disseminating written documents that summarize the evidence available on the effectiveness of different kinds of educational interventions.

According to its strategic plan, "the Substance Abuse and Mental Health Services Administration (SAMHSA) was created in 1992 as a result of the growing recognition that substance abuse prevention and treatment and mental health services are central to national health and productivity ( )." SAMHSA is now pioneering in the development of the Prevention Enhancement Protocols System (PEPS) to create program and intervention guidelines for the field of substance abuse prevention. Sponsored by the Center for Substance Abuse Prevention (CSAP), PEPS is part of the Prevention Technical Assistance and Training to the States (PTATS) project designed and operated by CSAP's Division of State and Community Systems Development (DSCSD).

Central to the development of PEPS guideline documents are the activities of a Planning Group composed of nationally recognized experts in substance abuse prevention research and practice. The Planning Group is responsible for reviewing the PEPS process, suggesting and prioritizing PEPS topics, and framing topic-related questions. The Planning Group is working in tandem with successive Expert Panels to establish "rules of evidence" for assessing practice and research findings and combining this evidence into prevention approaches.

The PEPS process begins with **topic selection**. Priority topics for guideline development are selected through consultation with a wide range of expert informants and analysis of documents relevant to the topic areas, including federal monographs, white papers, and annual data from substance abuse surveys. **Prospectus development** for each topic follows, outlining the proposed scope of the guideline. The Federal Resource Panel and Expert Panel review and revise the prospectus. PEPS staff then **identify research and practice information**, which is followed by assessment and synthesis of relevant findings. Three **documents** are then prepared for each guideline topic. *Guidelines* are monograph-length documents that provide a comprehensive analysis of the topic, full assessment of each prevention approach, and recommendations for implementation and evaluation. *Practitioner's Guides* are shorter versions that focus on effectiveness of the prevention approaches and recommendations for implementation and evaluation. *Informational Brochures* are intended for community groups and other lay readers to spark interest in the prevention topic. To insure acceptability as well as quality, all draft PEPS documents are reviewed by representatives from research, planning, and practice segments of the substance abuse prevention field. After **field review**, the documents are finalized for distribution to appropriate prevention audiences across the country.

ESTABLISHING EVIDENCE-BASED CENTERS: AGENCY FOR HEALTH CARE PREVENTION AND POLICY RESEARCH

Federal Agency	Agency for Health Care Prevention and Policy Research
Target population	Physicians, nurses and other health care providers, patients

Uses online information technology to identify best	No
Uses online information technology to disseminate best	No
Uses consensus/expert/screening panels	Pioneered in the use of consensus panels, now launching approach creating Evidence-Based Centers.
Uses rules of evidence	Yes. Evidence based centers "rules of evidence" for each prevention approach investigated.
Evaluation approach	Research and practice are carefully reviewed for specific approaches and then assessed for effectiveness.
Publishes hard copy resource materials	Yes.

KEY OBSERVATIONS

- The Agency for Health Care and Policy Research (AHCPR) does not differentiate promising and exemplary programs
- AHCPR pioneered in the use of consensus panels
- AHCPR is launching Evidence-Based Centers which review research and write reports that provide a basis for the development of health policy.

SAMHSA credits AHCPR with leading the effort to forge guideline development policy, procedures, and applications within the Public Health Service. AHCPR has pioneered in the use of consensus panels of both experts and lay public to develop clinical practice guidelines, releasing its first guideline in 1992. "Clinical practice guidelines have been systematically developed statements designed to help practitioners and health care consumers make decisions about appropriate care for specific health conditions. Guidelines have reflected current scientific knowledge of practices and expert clinical judgment on the best ways to prevent, diagnose, treat or manage diseases and disorders (AHCPR Fact Sheet, 1993)." AHCPR is now shifting to the use of evidence-based centers as an approach for the development of clinical guidelines. Evidence-Based Centers are awarded competitively to private contractors, who collect and review research evidence and write reports that provide the basis for the development of clinical guidelines. Evidence-based centers also organize review panels and oversee the development of the guidelines themselves (check this out). (Additional information will be provided on this process. What may be the first RFP for evidence-based centers is now being competed).

CONDUCTING CUSTOMER-ORIENTED EDUCATION EVALUATIONS: GENERAL ACCOUNTING OFFICE

Federal Agency	General Accounting Office
Target population	Educators
Uses online information technology to identify best	No
Uses online information technology to disseminate best	Indirectly. Online search engine includes full text of a number of publications that provide instruction in conducting evaluations. However, best practices are not identified by search engine.
Uses consensus/expert/screening panels	No
Uses rules of evidence	Indirectly.
Evaluation approach	Quantitative and qualitative measurement.
Publishes hard copy resource materials	Yes.

KEY OBSERVATIONS

- The General Accounting Office (GAO) differentiates successful practices and promising practices
- GAO conducts evaluations for Congress
- GAO develops and disseminate guidelines that assist others to conduct quantitative and qualitative evaluations

GAO APPROACH TO IDENTIFYING WHAT WORKS

The General Accounting Office (GAO) has recently conducted a consumer-oriented review of educational programs and practices for public schools (White and MacColl, 1995). The focus is both on successful and unsuccessful practice. Successful programs and practices are differentiated into two groups: (1) Successful practices are those documented by a body of literature, where outcomes are known and practices replicated. (2) Promising practices are those where initial outcomes may have been known but long term outcomes not, also practices not replicated widely. The customer for this review wanted to identify successful practices that could be implemented in a public school system. Criteria for success encompassed practices that include achievement, supported engagement, overcame disadvantage. The major conclusion was many practices can work but no one practice can assure success. The authors note that unsuccessful practices are often not described in the literature, although it is important to know about them. The authors also make the point that while the focus of their review is on innovation, existing practices may be also effective and should not always be abandoned in favor of what is new.

PROPOSING HEALTH CARE GUIDELINES FOR CONSUMER-ORIENTED EVALUATIONS OF HEALTH MANAGEMENT ORGANIZATIONS: AGENCY FOR HEALTH CARE POLICY AND RESEARCH AND SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION

Federal Agency	SAMHSA and AHCPR
Target population	Industry and business; individual health consumers
Uses online information technology to identify best	No
Uses online information technology to disseminate best	No.
Uses consensus/expert/screening panels	Consensus panels
Uses rules of evidence	No
Evaluation approach	Quantitative measurement.
Publishes hard copy resource materials	Yes.

KEY OBSERVATIONS

- The development of HEDIS (Health Plan Employer Data and Information Set) Guidelines has involved collaboration between AHCPR and SAMHSA
- HEDIS Guidelines were developed using experts panels.
- HEDIS provides a tool for evaluating HMOs in terms of performance indicators agreed on by experts

HEDIS GUIDELINES FOR IDENTIFYING THE BEST IN HEALTH MANAGEMENT ORGANIZATIONS

Fairly new, this is a consumer-oriented approach developed using consensus panels and focused on the evaluation of HMOs (Health Management Organizations) in terms of the quality of general medical care they provide and, more recently, the quality of mental health care (Mental Health HEDIS). HEDIS is supported and maintained by the National Committee for Quality Assurance (NCQA). HEDIS 3.0, the most recent version, was developed under the auspices of NCQA by the Committee on Performance Measurement, whose members have reflected the diversity of constituencies that performance measurement must serve, including purchasers (private and public, including Medicare and Medicaid), consumers, organized labor, medical providers, public health officials and health plans. Measures reflect a variety of performance domains, such as effectiveness of care, access/availability of care, satisfaction with the experience of care, cost of care. In each of these domains, the CPM sought measures that would be relevant to purchasers and consumers. Measures are required to meet criteria in three categories: (1) relevance, (2) scientific soundness (measures that are reproducible) and (3) feasibility (measures have to be easy to produce). Funding for the development of HEDIS has been provided by private and public sources, including AHCPR, SAMHSA and the Robert Woods Johnson Foundation. As a tool for evaluating competing HMO services in terms of a wide variety of specified indicators, it seems promising, especially if it should be made available to individual consumers as well as corporations and government. Extensive information about HEDIS is available over Internet. However, HEDIS is available from NCHQ and not over Internet. Information ranking HMOs is also not available over Internet.

OFFERING EXEMPLARY ARTS PROGRAMS FOR CHILDREN AT RISK: NATIONAL ENDOWMENT FOR THE ARTS

Federal Agency	National Endowment for the Arts
Target population	Educators
Uses online information technology to identify best	No
Uses online information technology to disseminate best	Yes
Uses consensus/expert/screening panels	Screening panel
Uses rules of evidence	No
Evaluation approach	Qualitative measurement, judgment of individuals trained in arts education.
Publishes hard copy resource materials	Yes.

KEY OBSERVATIONS

- The National Endowment for the Arts does not distinguish between promising and exemplary programs.
- NEA uses experts to evaluate claims.
- NEA guidelines are intended to provide consumers with a basis for making decisions about whether to replicate arts programs for schools and community.
- NEA disseminates publications evaluating the best in arts education over Internet.

NEA APPROACH TO IDENTIFYING WHAT WORKS

NEA offers numerous arts publications online, including publications for children. "Coming Up Taller: Arts and Humanities Programs for Children and Youth at Risk," available online, was prepared by the President's Committee on the Arts and the Humanities Programs for Children and Youth at Risk. "Coming Up Taller" identifies more than 200 individual profiles of programs that reach at-risk children and youth when they are not in school. In addition to the program profiles, "Coming Up Taller" provides the reader with an overview of the report and a brief history of why it was written. It reports conclusions, details the methodology, and outlines the organization. Recent research on the importance of arts education is also reported, including studies exploring the role of arts education in the development of higher order thinking skills, problem-solving ability and increased motivation to learn. Others studies, this publication reports, show correlations between arts education and improvements in academic performance and standardized test scores, increases in school attendance and decreases in school drop-out rates. Another publication that has been reported as available online from NEA is "Schools, Community and the Arts". This resource summarizes about 50 studies by persons who have knowledge of arts education, to be certain there are no exaggerated claims. Written in user friendly language, it is available in a number of formats including disk format.

CONDUCTING THE NAVY SCIENCE ADVISORS PROGRAM: NSAP AT THE OFFICE OF NAVAL RESEARCH

Federal Agency	National Science Advisors Program
Target population	Navy
Uses online information technology to identify best	No
Uses online information technology to disseminate best	No
Uses consensus/expert/screening panels	Possibly
Uses rules of evidence	Possibly
Evaluation approach	Mature technologies are tested in the field
Publishes hard copy resource materials	Yes.

KEY OBSERVATIONS

- The Naval Science Advisors Program (NSAP) conducts its own evaluations of mature technologies developed by the Office of Naval Research.
- NSAP sends science advisors onto ships to work with the Navy in testing technology.

NSAP APPROACH TO IDENTIFYING WHAT WORKS

Directed by Ms. Susan Bales, this unique program of the Office of Naval Research sends civilian science advisors onto ships to work side by side with Naval officers in implementing and evaluating mature technologies. Both scientists and the Navy benefit, scientists by developing first hand knowledge of the Navy's needs from the experience of being out with the fleet and the Navy through immediate and direct access to expertise provided by some of the best scientists in the nation. Naval officers select technologies of interest from a continuously updated Blue Book. Scientists who are planning to assume roles as science advisors receive several weeks of advance training that includes training in diversity (gender and race/ethnicity diversity training), with an assigned reading list that includes publications by Deborah Tannen ("You Just Don't Understand").

DISCUSSION

What lessons can be learned from the experiences of other federal agencies? As illustrated in this review, the federal agencies reviewed differ along a number of dimensions related to assessment of what works and to dissemination of information about successful practice:

- The Office of Adolescent Pregnancy Programs (OAPP) funds demonstration programs that are asked to provide rigorous quantitative evaluative evidence of their success. These programs are described to consumers along with the evaluative evidence but OAPP does not put a stamp of approval on some programs in preference to others in written documents, although program officers in discussion identify the best programs. Consumers make decisions as to the use to which they will put the evaluative information.

- For the Federal Emergency Management Agency (FEMA), a primary goal is making information available to states and local communities about emergency management preparedness practices that have succeeded at least once and that may be replicable elsewhere. FEMA does not, however, worry about failure of an approach described in its compendium of programs. By soliciting information globally, evaluating it and disseminating it, FEMA plays a unique role as a partner with states and communities in emergency management preparedness. Most of the time, it can be inferred, replicating a successful practice is better than trying an untested approach, reinventing the wheel, or taking no action. However, the emphasis on scientific rigor is lower than is the case with OAPP.

- The Agency for Health Care and Policy Research (AHCPR) plays a role similar to FEMA in its development of health policy (clinical guidelines) to guide both health professionals and the lay public based on research and practice. AHCPR has pioneered in the use of consensus panels and is now introducing Evidence-Based Centers that may play a similar role of collecting and reviewing research data, followed by writing of a report proposing health policy.

- The Substance Abuse and Mental Health Services Administration (SAMHSA) is using panels to help establish rules of evidence that can be used to assess success of substance abuse programs.

- The General Accounting Office (GAO) conducts evaluative reviews of research studies in a wide variety of federally funded program areas and also publishes guidelines to enable others to evaluate programs and practices.

- The Office of Naval Research's National Science Advisors Program (NSAP) program tests mature technologies by sending science advisors into the field to work with Navy crews to test approaches.

OERI may find it helpful to borrow components of all of these approaches:

- Like OAPP, OERI can call for rigorous evaluation in terms of a core set of indices common to many different educational programs, whatever their differences. Such information can then be reported to consumers.

- Like FEMA, OERI can call for nominations of exemplary practice using advanced information technology on Internet and disseminate resource guides to the educational community using Internet as well. FEMA's approach seems particularly well suited to OERI's information clearinghouses .

- Like AHCPR and SAMHSA, OERI can use a system of expert panels and in fact is already doing so. OERI may also find AHCPR's latest efforts to create Evidence-Based Centers that collect and review evidence and publish reports and SAMHSA's rules of evidence meaningful models for the development of educational policy as well for the development of criteria against which the performance of educational programs and practices can be assessed.

- Like ONR's NSAP program, OERI might consider sending educational advisors drawn from universities and educational laboratories into the field to assist local school districts and communities in ways that states may not be able to do as readily.

Are there lessons also to be learned from a review of the approaches different agencies take to the development of performance indicators? Identifying performance indicators that are both measurable and meaningful can be challenging. OAPP appears to rely principally on a small number of performance indicators common to all adolescent pregnancy programs, especially, birth weight, incidence of repeat pregnancy and age at first onset of sexual activity to assess the success of the programs. As important as these indicators are, there are other kinds of evidence not as readily measured which OAPP-funded programs do not report, such as early childhood development of children born to adolescents, educational outcomes for the adolescent parents, employment, marital history, substance abuse, etc.. The picture that emerges of the program is partial at best. Moreover, even health indicators such as those used by OAPP are not necessarily free of cultural bias, although they would seem to be less likely to be influenced by cultural factors than, for example, standardized intelligence tests. Some racial/cultural groups (e.g., the Hmong) encourage childbirth not long after a girl is capable of reproduction and the fact that a program for these populations does not succeed in reducing the incidence of adolescent childbirth may not mean that it is not a successful program - it may be very successful in achieving its other objectives. It may be helping, for example, adolescents become loving parents.

More broadly, it seems clear from the examples provided by these agencies that as important as quantitative measures are, they should not be the only indices in terms of which programs are evaluated. We are often too hard on programs, rejecting them when they fail to meet minimum standards of statistical significance even though qualitative measures — one on one interviews or focus groups — would reveal that they have had important effects. OAPP has noted this in conjunction with its study of Adolescent Voices. Teen mothers are reporting successful outcomes of adolescent pregnancy programs that do not produce statistically significant results. FEMA asks that a program have succeeded at least once — in the perception of those in a position to make an expert judgment, and not necessarily statistically. AHCPR has pioneered in the use of meta-analysis, in which the results of individual studies that do not reach statistical significance are pooled over multiple studies. By extrapolation, OERI may perhaps learn from these examples that quantitative approaches should not be the only approaches in terms of which educational programs are assessed and that statistical significance should not constitute the only criterion in terms of which programs are judged. The failure to reach statistical significance, in other words, may have more to do with the imprecision of our measurement instruments than with failure of programs to produce an effect.

One qualitative approach this evaluator/reviewer recommends for use in collaboration with a quantitative approach involves the critical incident technique. The critical incident technique involves both qualitative and quantitative components. It can be used in both formative and summative evaluation. Originally developed by Flanagan (1954) in the 1940s for use in training programs being designed for the Army, the critical incident technique has been used extensively to improve training programs. Critical incidents are best and worst, or most and least successful, experiences that people report in interviews or self report surveys -- for example, as applied to adolescent pregnancy programs, they can be the behaviors or skills that keep adolescent parents in school and enable them to graduate. When reported by students or their teachers or even their parents, critical incidents can call attention to the dynamics behind the numbers. It can help explain why an educational or training program works or doesn't work.

This reviewer has also found it helpful to use an approach which can be described as a meta-evaluation to look for effects in a group of programs that may not be apparent if they are examined one at a time. Like meta-analysis, this approach involves looking at effects across many different programs. Unlike meta-analysis, however, the approach looks for critical dimensions that differentiate the clearly most successful programs (based on earlier formative and/or summative evaluation) from other programs. This approach can be especially helpful in calling attention to critical aspects of programs that were not recognized as important when the programs were evaluated one at a time. For example, in an evaluation of 20 college/community partnerships, in which colleges teamed with their local communities to create college-bound programs for economically disadvantaged multiethnic precollege youth, a number of programs succeeded beyond all their expectations. On the other hand, some did not succeed at all. The meta-evaluation showed that the best programs provided opportunities for students to *bond* by bringing them together in ways that enabled them to share intense, positive experiences; to develop a new peer group consisting of other student members of the CCPP; and to establish a common set of norms or values, most especially, recognition of the value of education. Fortified by these experiences -- summer camps together, overnights on college campuses during the academic year, out of town trips to college campuses and cultural events, Saturday academic sessions -- students could more easily cope with pressures from other peers, in school and on the street, who did not have these values. Less successful programs approached students one at a time only, for example, in the form of one on one mentoring or tutoring in their classes at school.

If there are lessons to be learned from other federal agencies, there are also lessons to be learned from OERI's example of a two-stage evaluation process that differentiates promising from exemplary programs. Consumers may do better replicating programs that have been shown to work in multiple settings, all other things being equal. On the other hand, it is probably better to replicate a promising approach that has been shown to work in one setting than to take no action at all.

At the same time, Inherent in the notion of promising and exemplary programs is the concept of innovation. Promising and exemplary programs are so designated not only because they are successful but because they represent new approaches that are better than the old approaches used in the past — whether by reducing the incidence of low birth weight in babies born to adolescent mothers, increasing emergency management preparedness, improving health care, reducing substance abuse, keeping the arts in schools, or in any of myriad other ways. Not only can promising and exemplary programs, policies, practices be replicated in other sites, but they also encourage further creativity in addressing problems that seem to defy solution. Educational, social, environmental and health programs can only benefit from efforts to identify and disseminate the best.

BIBLIOGRAPHY

Adolescent Health Report. Rockville, MD: Maternal and Child Health Bureau, Health Resources and Services Administration (1996).

Adolescent Pregnancy Prevention: A Compendium of Programs. Washington, DC: Healthy Mothers, Healthy Babies Coalition (1995).

"A STEP-UP for Arvada's Teens". Arvada, CO: Jefferson County Department of Health and Environment.

Coming Up Taller: Arts and Humanities Programs for Children and Youth at Risk. Washington, DC: President's Committee on the Arts and the Humanities Programs for Children and Youth at Risk (1996)..

Data on number of teen mothers giving birth every year. Washington, DC: *The Washington Post Parade*, January 1997

Preparedness. A Compendium of Exemplary Practices in Emergency Management. Federal Emergency Management Agency (1996).

School-Based Programs for Adolescent Parents and Their Young Children: Guidelines for Quality and Best Practice. Bala Cynwood, PA: Center for Assessment and Policy Development (1996).

School-Based Programs for Adolescent Parents and their Young Children: Overcoming Barriers and Challenges to Implementing Comprehensive School-Based Services. Bala Cynwood, PA: Center for Assessment and Policy Development (1994).

Schools, Community and the Arts. Washington, DC: National Endowment for the Arts.

The Adolescent Family Life Demonstration Projects: Program and Evaluation Summaries. Bethesda, MD: Office of Adolescent Pregnancy Programs, Department of Health and Human Services (1990).

G

NOTES ON THE FEDERAL DISSEMINATION EXPERTS MEETING:
"Advising the Department of Education's Office of Educational Research and Improvement on its Design of a Systematic Consumer-oriented Evaluation System to Designate Promising and Exemplary Educational Products, Programs and Practices"
Oct. 12, 1995, Academy for Educational Development, Washington, DC.

Karen Bogart, PhD
Anne Steinmann Institute

OERI PERSONNEL

Reauthorization legislation - new ideas about how to do a better job of disseminating the best for educational improvement.

Lawrence Peters, OERI. "Legislative Background on this new dissemination and evaluation challenge". (Peters was formerly Counsel to the Select Education Subcommittee headed by Major Owens who was the primary author of the OERI reauthorization legislation.

Lawrence Peters, OERI. Federal Education Research and Development is underappreciated in terms of budget and in terms of the way the public sees its operation. Like Internet, it is a great innovation. But a cynic replied all the books were on the floor. What is needed is a sense of how to get all the riches, fabulous wealth, out of resource - Analogy appropriate to OERI. Maybe books not all on the floor but all in closets. Inaccessible. Hard to get to. National Dissemination Network - NDN - approves exemplary curriculum. Panel of people approve. And then money is available for schools interested in replicating that particular innovation. Marvelous system and it has made a great impact on schools throughout US. Problem was that many of the programs that were exemplary or could be so classified could not get through this enormous pipeline to be approved by this distinguished board of people. There is a sense of a lot of stuff on shelves growing stale and not a whole lot of flow through. ERIC - overstuffed closet - everything was there - any literature at all could make it through but it was difficult to ascertain what was quality and what wasn't. Our view - how to try to determine a way in which more good stuff - exemplary curriculums, exemplary innovations - could be reviewed at a national level and given some kind of status that would make sure more people around the nation could take advantage of these resources.

Susan Klein made a lot of contribution to the initial legislation - she had looked with colleagues at NIH and its use of consensus panels - as a way in which people get together and make decisions about controversial issues and come down with agreements that helped the profession, in this case, medical profession - consensus panel idea appealed to us in that we were skeptical of idea that defense had all the wisdom about ed practices and programs to be implemented by practitioners - the notion came up of modifying consensus panels to incorporate practitioner knowledge in making judgments about particular educational practices. That fitted into our notion - if OERI was to see increase in respect it received, there needed to be stronger ownership by practitioners of the kinds of practices and information that was being disseminated. We are struggling with notion of how to get right balance of people who can make right judgments of people who can make decisions about educational practices and to use the rest of the system - to unlock closet - and make sure more programs are reviewed at this level.

Eve Bither, OERI. "The Role of Consumer-Oriented Evaluation in the National Education Dissemination System" Eve wants to describe within the context that Lawrence has set for us - the new legislation and priorities -one of the activities assigned to OERI and to its new board - that just had 3rd meeting. One of the assignments for OERI and Board talks about establishing standards for a variety of activities that had always been a part of work of OERI. The first set of standards to be developed were standards for selection of grants. Standards described procedures and people who need to be involved in selection of discretionary grants for all programs in the Department. Completed recently. Two other parallel efforts - one developing internal guidance on how they will monitor the grants that will be awarded in future. Internal guidelines. At same time, own office will be developing a new laboratory system for the laboratories - which are contracts. New RFP calls for a new way of office to interact with regional laboratories. Third - mandate to us to develop standards for identification and validation of promising and exemplary practices. While NDN is only program within the department that uses clearly defined standards and criteria and an expert panel for declaring something to be validated, there is a continuing search for including not only exemplary practices but promising practices - especially when educational change is rapid, when new technologies are appearing. Conference in Feb that they held with NSF - one frequent request was a cry from participants who said NSF sponsors a great number of curriculum development projects, many teacher enhancement projects - which of all of these things are really good. Are 16 curriculum frameworks all equally good? The standard setting process is a response to those requests. Currently some doubts about whether there is a federal role in education but area of identifying and recommending exemplary practices is recognized as an appropriate role for government in education - no debate about that. Want from all of you suggestions about proposed standards. Want comments on the criteria we have proposed and on the process that we propose in implementing. Draft - standards for the identification of exemplary and promising programs. Use of word program. Most people in education think of a curriculum program or educational program - includes in legislation everything from research program to curriculum design to policies to educational practices. How would you make it possible for programs to become a known to make it into system as exemplary pr practices.

We propose to make it as inclusive as possible. We also want to make it possible for panels or OERI to accept submission from a wide variety of developers and providers of such information. Many reform networks, many states, also have some outstanding ideas that should be shared with nation as a whole through this process. Want also to validate this information through expert panels - want to establish a pilot panel in Math and Science jointly with NSF. Also help develop process by which they will use pilot panels in general. In addition, want to charter other panels to validate the decisions of other federal agencies and others in identifying promising practices - should represent researchers and practitioners, and experts in evaluation, and people familiar with educational reform in general. Also issue of length of time for which something would be validated -- our board definite about fact they did not want a definite time period. Rather a date at which designation was made and leave it to the judgment of potential consumers to make conclusions from that.

Finally, criteria for standards, for designating programs as exemplary and promising - need, significance, quality, effectiveness, usefulness to others, context. We think equity issues need to be incorporated in each of these categories of criteria. But in general we have not had any negative comments on what we have included or what we have left out. Purpose for doing all this is to have worth while information about things at schools and that teachers and administrators could do to improve education for students - for National Educational Dissemination System.

Sue Klein, OERI. "How we hope to learn from other federal agency experiences." Sue wants to emphasize that we are at the beginning stages of the process. One purpose of having you here is to learn how you are approaching this process yourself. One term we are playing with is - "choice-oriented evaluation" - we need to figure out some ways to differentiate this evaluation function from other evaluation functions. Wants to introduce key players. Capture good practice from federal agencies outside education. In addition to math and science panel, folks from Education Development Center are here. They are operating pilot expert panel in gender equity through their Women's Educational Equity Act Resource Center. Another special participant is Eleanor Chelimsky, recently retired director of the U.S. General Accounting Office, Program Evaluation and Methodology Division. She will talk about relationship between evaluation and dissemination.

Eve Bither noted that we are at the very beginning of our standard setting process. We have until end of March to come up with an acceptable set. Ample opportunity to hear from all of you.

PANEL OF FEDERAL AGENCY EXPERTS IN CONSUMER-ORIENTED EVALUATION SYSTEMS

Conrad Katzenmeyer, National Science Foundation. Conrad is a former colleague from OERI. He is now representing NSF. NSF does not have auspicious history of doing this kind of thing. We do an excellent job of reviewing content - this is hallmark of our effort around curriculum - it is a very good process. We do not have panels looking across various curriculum development materials in general - we did it last in 1976. That was end of curriculum development era in federal agencies. It was strictly a content review. Panelists looked at materials that were developed,. Not much data beyond that. Decisions made. Some terminated, some enhanced. At least some history of having done that at the time. We are now embarking again at a panel effort around our curriculum development materials at k-12. For half million dollars, we will have panel that will review 25 of our programs and 15 others - contractor will collect information in regard to those curriculum development efforts - evaluation and dissemination. We hope at end of 18 months to have some reading of how those particular programs have fared. What we have learned - or I have learned - from time on program effectiveness panel is panel can do a very good job on content review, can do a very good job on reviewing hard data on piece of paper - it has great difficulty beyond that. If we were to limit to content review or do what JDRP or PEP has done in past, we probably would not do it at all.

Can they give leads on methodology on how people should go about evaluating - did describe at AERA.

Kathleen White and Gail MacColl, U.S. General Accounting Office. Gail wants to share observations she encountered on their report. She works for GAO. Customer is Congress-- The House Subcommittee of Committee on Educational Opportunities and they asked us to report to them on what works and what doesn't both in schools and the workplace. This request involved a formidable search of the literature to identify relevant studies we could review. Had to scope down to be manageable. Scope down research - solid research, well designed, with measurable outcomes - distinction between research that documented successful practice, research that documented promising practices. Successful practices - documented by a body of literature, where outcomes were known, practices replicated. Promising practices - initial outcomes may have been known but long term outcomes not, also practices not replicated widely. Customer wanted successful practices that could be implemented in a public school system. Criteria for success - practices that include achievement, supported engagement, overcame disadvantage. Major conclusion was many practices can work but on no one practice can assure success. Gestalt that work. Some additional observations about the literature - very few accessible and user friendly summaries of successful school and work practices, many evaluations were difficult for layman or even teachers to understand because of jargon, sufficient evidence to document practice very rarely emerged from one single study - most from multiple sources using a variety of methodologies.

Kathleen White, second GAO speaker. They had six weeks on what works and six weeks on what does not work. Largely an annotated bibliography. Includes caveats on what was covered and what was not covered. What may get evaluated may be effects of a practice on narrow sense and not effect on larger classroom. She wants to make comments about methodology that occurred to them as a result of having done this study. There is a downside to focusing on what's new and what's best. Both in schools and workplaces, they can suffer from always adopting a latest fad and therefore never sticking with any practice long enough to be effective - dysfunctional. Second - consumers in schools are reasonably skeptical about things being effective - great claims for things but not effective. Third - information about best practices does not necessarily give consumers all information they need - only practices that get evaluated are ones people nominate - advertisements for packages can sound wonderful. Three questions - what about ordinary or conventional practice - what if we find successful schools. Some of schools designated most successful used what researchers describe as ordinarily teaching, not cutting edge. Schools that really are failing, what they have is no teaching. We can forget that ordinary practice well done can produce very good results. We don't want to abandon conventional practice. Secondly - new practices go in cycles and often inappropriately discard strengths of what came before. We don't want to abandon what is old in our concern for what is new. Practices grounded in research - let us not forget practices grounded in social science knowledge. What doesn't work. Our committee wanted to know as much about what doesn't work as what does. Difficult question to address. We did not find much on what does not work - e.,g like individual merit pay for teachers. But clearly what doesn't work is equally important. From the consumers point of view they need that information as much as they need information about what is best practice. We did bring another report - Program Evaluation - organized on what is most useful. It has a handy list of key questions for evaluation you might want to ask for any program. Also discusses how audience knowledge of what questions to ask - the audience in this case being Congress.

Linda Blackenbacker, Agency for Health Care Policy and Research (AHCPR). I am not an expert in guideline dissemination. But I have been working with consensus panels of experts. I have worked with intramural and extramural researchers. Treatment information in oncology. Worked at NIH Consensus Development Program - convene extramural panels for evaluation. Not at all involved currently in development of clinical process guidelines currently. She is involved in peer review process to assess impact of information of federally disseminated guidelines. Looks at behavior change in the clinical community - consumers being all kinds of people. Looking at how we provide information and assess it. That is context for the way I will tell you about this agency that is authorizing legislation for the preparation of clinical practice guidelines. Just issued 17th guideline on cardiac rehabilitation. Our publications catalogue lists most of them. Let me tell you how we do it. Essentially it is a rigorous process compared to those of NIH or American College of Physicians also doing this. Process we undertake is the most rigorous and also fairly costly - $600,000 top $1 million per guideline. How we pick topics - they represent health problems as opposed to use of a technology - can be a specific diagnosis such as nonfatal angina or a condition that would cut across. One of earliest guidelines - was Guidelines address very complex health problems. Multiple factors guidelines have to consider - onset, procedures, consequences of procedures with respect to diagnosis, treatment, prevention. Labor intensive process panel members go through assessing the literature. Because so much expertise is brought to bear on this, fair amount of labor they have - multidisciplinary - can't be just internal medicine, have to cover full range of disciplines. A lot of knowledge brought to panel through widespread review of literature. Some funding from Medicare trust fund. Some initial topics came because conditions prevalent in Medicare population. Broad range of persons interested in issues. Panel chair and co chair. Detailed specific policies for selection of chair are announced in the federal register. The announcement says they are seeking to establish a panel. They contact professional groups and consumer groups. Criteria for selection on panel - relevant training and clinical expertise, some prior experience in guideline activity, person recognized by community for his or her contributions, demonstrated capacity to lead persons from different perspectives, some one who is fair and objective and can leave biases at the door. Have to be recognized for fairness and objectivity. Panel members send in names to agency for consideration. Panel has to be workable size, not more than 20, the larger the panel the more multidisciplinary - there has to be an epidemiologist, health sciences researcher, and social scientist - not just a bunch of doctors. Guidelines panels has to have a consumer representative.

NIH consensus panel also has to have a consumer rep. Balance on the committee, not a small cadre of folks who will overwhelm other bunch. Agency administrator issues the formal invitation. very specific rule s and procedures for going through the guideline development process. Some publications - methodological perspectives - clinical process guidelines. Includes metaanalysis. Process for producing guidelines varies.

Herman Disenhouse. Substance Abuse and Mental Health Services Administration (SAMHSA) Joined SAMHSA in working for Treatment and Improvement in 1989 coming from Institute of Medicine where he worked with other types of consensus panels. One task was to look at NIH Consensus development Model, to look at AHCPR which was going into its practice guideline. One of things we did was develop five or six different strategies. SAMSA was split off from ADAMHA about 3 and one half years ago when research arms - NIAA, NIDA, etc were sent back to NIH. We were supposed to be services funding and dissemination. Dilemma within NIH - when you hold up consensus development panel and practice guideline panel - recognize we have our fads and fancies in R&D and D&R - much of our emphasis focused on D and not R. One of things ED has to do - when we started we were told if we did what AHCPR did we would have to publish them as regulations- 2 or 3 million dollars - which we did not want to do - so we invented treatment improvement protocols which are consensus panels that are brought together to review the literature, sometimes including metaanalysis. Their customer is Congress where the standards of evidence are not as high. So we need to recognize that there are at least 7 or 8 standards of evidence out there - very active public information dissemination and professional dissemination - there are now 17 treatment guidelines published and no funds to republish - wanted to evaluate them. Anecdote - this year he conducted 2 review panels - in one of the study sections, reviewing applications for replications - 5 replication projects - chosen by a peer review panel of evaluators - applicants could apply for a self replication or another replication - review committee repudiated findings of earlier panel, did not agree that these projects were exemplary. Also did a Criminal Justice Treatment Networks Demonstrations Projects - one of the applicants proposed a methodology that had never been tested and was in conflict with other methodologies - question as to how to evaluate methodologies.

Ron Smith and Joan Dilonardo, SAMHSA. Develop evaluation information system for SAMSA. They brought examples of the kinds of things they are doing to evaluate. Some are directly result of consensus process, other things have an expert panel or consensus panel built in. This year to develop the focal topics - took agency and people in field on a retreat to develop the topics for evaluation. In addition, SAMSA has an evaluation policy throughout the agency that has methods for evaluation in the demonstration grant. Includes local evaluation in demonstration grant. That was beginning of process. Middle. Before Consensus Panel, Resource Panel that identifies people - federal agencies, professional organizations, etc. - job is to figure out is the field ready to write one of these - consensus guidelines. Also approach looking at different approaches in the field. Also example of report they produced which is an assimilation of 25 different studies. Also have done a bibliography in terms of training for treatment providers. These are kinds of things they are trying to pull together in terms of the development of their standards.

Michael Sikes, National Endowment for the Arts (NEA) We have had our share of publicity. Within NEA is Arts and Education Program for which he is the Assistant Director. Small portion of the budget of the total Endowment. Seven million dollars to fund two initiatives - Category - to State Agencies. Other a grant to an arts organization to develop a partnership with a school or schools to fund programs. Arts and Education program will cease to exist end of this year. Moving away from 15 different disciplines. To what extent K-12 education will continue to be a focus is anyone's guess. Not likely. All federal agencies have had to undergo cuts. What we have tried to do in the past, our biggest challenge has been to put the arts back where they belong. But to go back to what we did do, the Endowment has used the system of expert panels to validate and assess proposals, to make awards - State Partnership Category and -. Panels have consisted of arts educators, general educators and laymen. Used very effectively. We have tried to emphasize systemic educational efforts working in collaboration with state departments of education. We have felt that in order for arts to reclaim their rightful place among pantheon of core subjects we need to focus on arts as being part of a systemic sequential curriculum. He has been aware of criticisms of The Endowment - whole notion of federal funding of the arts. Over last year, change toward idea of marketing - the kinds of processes we use in the Endowment to bring in expert panels - proper way of conducting an agency program. Problem has been not marketing that expertise in the way people can understand. Article in NY Times criticizing the whole panel process - people giving grants to other people based on insider relationships - and those of you who work with panels know that is not the case. Main point - in trying to assess what we might do with limited research funds a year ago, we decided to bring together an existing body - and the result is a document called "Schools, Community and the Arts". A few characteristics of it - it summarizes about 50 studies by persons who have knowledge of arts education to make certain there were no exaggerated claims., It is written in very user friendly language. It is available in a number of formats including disk format to put it in a computer, also available on the World Wide Web. We wanted to make it as usable as possible, a lot of people are not aware there is good research about the arts. If want a copy - 202-416-? Web Address - HTTP://ASPIN.ASU.EDU/ rescomp.

Vernon Adler, Maria Mlinarcik and Len Oberlander, Federal Management Agency, FEMA. Vernon Adler. About a year ago, the Federal Emergency Management Agency faced the fact that they were going to have to live with, make do, and states in turn in field of emergency management, if we were lucky, with as much money as we had from year to year, and more likely less to do the emergency job in this nation than we had. So we set about a task of how to leverage existing and shrinking resources in this field. It emerged as partnerships in preparedness initiative. The initiative began with the letter from the Associate Director for our Directorate - Kay Goss - to all of our 10 federal regions to solicit promising or exemplary practices. We are using words interchangeably because this is becoming a jargon melting pot for same goals we are all espousing. I want to show view graphs on exemplary practices in emergency management to a reality. The initiative is an agency initiative. "Preparedness. A Compendium of Exemplary Practices in Emergency Management". Our partners are State and

Local Governments - supported through a flow of funds through Congress to FEMA and on to the States through Local Governments. Our job is to build the strong and effective emergency management system and build on the system that already exists. We continue to search for creative ways to use limited resources. That is what we have been talking about all morning. The subject here is emergency management. We could talk about specific goals of a dozen or more federal agencies. What is an exemplary practice in our definition - a program or project or technique that works in one place and is adaptable by others - somewhat subjective, rather open ended. One limitation - it must have worked. So we get into the question of how to judge that it has worked and is adaptable. It comes from our constituencies that are not just local agencies - State and Local Govt, business, Chamber of Commerce - with good ideas they have gotten into practice, so they have an empiricism to report on their practices. This Compendium will describe the public and private sector initiatives and does so - first compendium is approximately 100 practices around the country that have been screened and work. The Compendium will refer the individuals to the practices - points of contact - that have made practices a successful reality. It is most important we communicate information effectively, so compendium is a vehicle - we are also using for solicitation of practices WWW. It empowers the Emergency Management community to take existing talent, existing resources, and leverage it. A simple statement. Let's keep it simple. The philosophical, high minded, erudite way in which we talk about our problems and sometimes lose our audience -

Maria Milarcik. FEMA. I want to say a few things about the development of the Compendium of Exemplary Practices. We are surfacing new ideas through this initiative - impetus for it is a lot is happening in the private sector, at the local level, in businesses, community groups. We want to surface these things as well as other successful ideas and programs. This is not a program for policy formulation at the federal level. We communicate directly with local practitioners, with individuals. We do not evaluate practices. The screening panel screens practices to assure that definition of exemplary practice is met. Practice that has worked in one place. We are not concerned with how efficient it is, how effective, we leave that to individuals who use it. We are fostering partnerships and communication at all levels. The screening panel passes on nominations, they do not judge quality, significance, need for this - that is up to users for them to decide for themselves how they need it -- how it may be significant to them and, how they may want to adapt it. We don't promote it, we bring it to the attention of practitioners for them to assess it relative to their own needs and resources. We did not spend any money on program development. Money came out to $20,000 per production of Compendium. Reason so efficient it does not cost anything to put word out through WWW, or to receive responses. Only cost is cost of production and printing. We have invited participation in this of everyone in emergency management. Partnership Preparedness - book - can't give it out because not yet approved. Scheduled for publication in December. Format selected was designed to be very user friendly so anyone skimming through could obtain the information for their own use. Also inclusion of four indexes in back of book - title, location, point of contact. It will be available on WWW.

Discussion with ED Staff asking Panel Members Questions

Sue Klein: OERI will talk about system we are trying to design.
Discussion. ED Panel members come up

Questioner: Elizabeth Payer, OERI. I am posing questions as relative newcomer to evaluation. What were some of the key practical and political issues you faced in your review processes? Question - dynamics and relationships in all organizations - how did this affect what you did.

Discussion:

Herman Diesenhaus (SAMHSA). The only thing I could think of - the most widely disseminated education prevention effort is Project DARE - it is in every school in the country - and all evaluation data says it has no effect. I believe we have accomplished a lot of good with this program and I don't believe that it has not been good. It has gotten cops into schools, reduced social distance between cops and kids. To say it's a failure on some goals may be true, but as a social experiment it is not a failure - it was disseminated as an exemplary program and became a practice guideline. I was not always a fed, a fed when started, left, came back. In every setting I have been I have not seen it work the way we want it to work.

FEMA representative (Len Oberlander). The context in which we answer this first question. Motivation. Expectation. We are talking about different constituencies. In the emergency management community, when you see and read about response to hurricanes and suffering of citizenry, effectiveness of that emergency management is measurable quantitatively, qualitatively, subjectively and it all has to do with perception of audience - citizen - person being helped, person who did not get enough help, person reading about it. Complex. In our business, screening panel is a cross section of experts in the field, that is a must. I heard earlier about 20 or more people being on a screening panel, I heard that they have a lot to do with setting up standards. It is a problem when the federal agency is seen to be prescriptive with the users. There is a lot of public education associated with the evaluative process, it must be user friendly, when you have 20 or more people, blue ribbon, there is a danger of distancing from the user - person who walks into hospital, parent who sends child to school, person who lost home in mud slide - you are dealing with explosive emotions and a variety of perceptions. In terms of emergency management, we must be humble and sensitive to what our constituencies believe us to be in Washington. I want to interpose those views because of the selection of our screening panel - I don't like the word - initiative in our business - if it is a first responder like a policeman, etc - that recognition or feeling of pride could be a certificate or an award, does not have toe money, it has to be a recognition of that person's doing for the community. Herman's point. Goodness has to be goodness for what. Who is the customer? The spin off of that program was particularly good for a particular purpose but was not the stated purpose of the program. Possibly the outcome you are measuring is wrong.

Herman Disenhouse - focus of issue I hear being discussed - we have different effects of federal agencies calling something an exemplary program.

FEMA. Good way to frame the issue is what is the effect of the federal government calling something promising or exemplary 1- as in FEMA, information is being shared and anyone has access to that information and can use it as they choose, so in own home, school, fire station or whatever they can decide what significance that practice has to them, whether they need it. 2 - the other side is when something is called promising and at the federal level an entity takes that and produces something prescriptive, such as a guideline and attaches that as a criteria for grants - you are taking away individual ability to use that information the way individual wants. Should federal government take away individual ability to use info at local level. Who is the user of the information we are trying to deal with? Evidence doctors will ask for will require quite a large panel - under what circumstances is it true or not true? Other situations. Why do we all have to do it the same way? Important who the user is.

Herman Diesenhaus - To me a guideline is not prescriptive. For 10 years I was responsible for licensing of (medical?) practices. Our tips are not standards, they are guidelines. Eleanor Chelimsky is right.
Different user groups look at it differently. You may list 50 exemplary programs, 10 may be adaptable to a variety of situations, others not. Many people don't make the distinction. I get phone calls where people treat our tips as regulations.

Linda Blackenbacker, AHCPR - Problem with our - AHCPR - production of clinical guidelines is medical population is so volatile. Pocket book issue that emerges from production of guideline. We market distribution but we don't market sufficiently well that this is not a federal government pronouncement - this is an independent panel that looks at the literature - the notion is we are prescribing, but we are facilitating assessment of information. We don't do a good enough job of saying it is an independent panel doing this.
GAO, Gail MacColl. One risk is not only fed govt. but also state supt or local supt may say this is so great we will put it in our schools - may be secret to success of method if it was developed by teachers in a particular school and it was their enthusiasm that was secret to success, and you don't find same result when you try it elsewhere. When we were trying to settle in our own minds, what do we mean when we say something doesn't work, is someone can say, oh, I know it worked in x community - they are saying, most of the time it doesn't work even if there is an isolated success.

Questioner: Cindy Stewart, Office of General Counsel, ED Question on legal issues. Technical credibility question. Exemplary in terms of imprimatur - we will be assisting OERI in developing standards.
What kinds of concerns should Dept be sensitive to in terms of setting up procedures? We would welcome comments you may have. How do we ensure objectively and lack of bias on part of experts and practitioners?

Linda Blackenbacker, AHCPR. Guidelines panels members have to fill out financial disclosure form, demonstrate they are not in conflict of interest - wont benefit financially. We put those in the federal register.

Herman Diesenhaus, SAMHSA. Very important to differentiate between issues where federal govt. has a role in protecting the public interest, as in medical area, where people need to be protected because they don't have ability to make judgments in area, and that area has to be separated, differentiated, from area where knowledge may reside locally. Process of seeking comments and resolutions is extremely important where federal government has protective role and in other area individual citizens and groups have to see federal register and follow it.

Federal Register for a lot of folks is unknown and irrelevant - how many school principals know what it is.

Herman Disenhouse.. We have a block grant . We sent copies of Federal Register to everyone who might have interest in the topic and might not see the Federal Register. The more you can do to outreach in getting comments the more you can avoid the pitfalls. Every state has an administrative procedures act - you are codifying behaviors in education much more than in health when you use procedures like this. I don't feel all knowledge is in federal government. Our jobs as feds is in bringing it together., If we want to reach practitioners, publication is going to have to be done in the content and grade level journals so that teachers actually get in their hands publications that they read regularly. I don't enjoy reading the Fed Reg - summarizing it is a good way to go about it. Credibility issues. Outcomes. At different times, people have said in different ways we want to focus attention on programs that work. Issue I want to raise, in context of education, is what does that mean? In education, we work in area of research and improvement - public perception in country that level of achievement is not satisfactory. To press upon us is to fix that, to
come up with solutions. It turns out not to be easy to do. Notion of expert panels in area of education is to find some treatment or package that shows that student achievement went up. Counterparts to that in other areas we are talking about - quicker recovery from illness, better response to emergency condition, lowered drug use. Two questions. If you agree, with siege mentality, to ask about your context, is that something you are operating in, or do you perceive yourself as in a private company, focusing on increasing profit margin? And if so, to what extent do you think we should be g trying to conceive of innovations that will bring about demonstrable achievements?

Michael Sikes, National Endowment for the Arts - In the Endowment for Arts, we are in same situation as OERI. We are trying to market educational achievement but we are trying to do it outside hubris OERI would do it in. In Endowment, 1990, sponsored a study on how arts contributed to excellent education - charge to find programs that contributed to general education through arts infusion using criteria such as
standardized tests - what they found, drawing on people like Howard Gardener, is that is wrong approach, we cant make best test for arts on SATs, that we are not capable of making case for

arts education based on what it does for arts and science, yet that is what people want. We need to reeducate the American Public as to what education involves - new theories of learning such as constructivism, that tell us learning
is an extraordinarily complex process.

Conrad Katzenmeyer, National Science Foundation. - Pressure for us to be responsive to student achievement as an outcome which we resist like crazy. Yet we have strong culture in agency to that says it is total nonsense to ask that question and what is where I see panel playing an important role - panel will be sensitive to broader range of outcomes than student achievement.

Linda Blackenbacker, AHCPR. Our guideline development is a companion piece to our larger program which is to look at what works. However, much of medicine is practiced in absence of a lot of quantifiable evidence and the guideline panels grapple with lack of documented evidence. Like to base good decisions on randomized control trial. But our goal is to look at what works and what are effects on patients.

SAMHSA. Our agency always looked at public sector as our client but in health care you can't. You must have sitting at table private practitioners. Now health care industry is developing HEDIS data set and AHCPR plays key role in funding this - we look at organizational format. Be sure you have representatives of those organizations on your panel that you might not have if you only had a public sector mentality.

FEMA. Compendium is in its infancy and book coming out, on Internet, is first of its kind in emergency management infrastructure. What I must do and what FEMA must do is an outcomes management survey and we need to look at feedback devices. When info is disseminated, we must set up system whereby people who take from compendium and adapt to their own needs, we must get them to feedback to us, and thereby make continuing assessment of the product. That will be a reality check. Not only reality check but one we have to undertake.

GAO . It is useful for us to distinguish between curriculum and instructional practices. Schools doing worst are dysfunctional institutions and no curriculum is going to solve that problem. Different situation than in a middle class school that wants to know is this the best science curriculum. Engagement. A lot of lit on effective schools says if teacher and students are engaged in their work, more effective than if not. Best schools - where kids come to school, are excited there, are participating - that is a dimension parents are looking for too. You don't want to see just test scores, but it is it an exciting place for students to be.

Questioner: Carol Chelemer, OERI on Management Feasibility Issues - Management Capability. When we are thinking in OERI of exemplary practices, we want to cast a wide net, some suggestions on solicitation processes. We don't want panel to be overwhelmed. How do we cast a wide net, yet use some screening mechanisms to make work of the panels screenable?

FEMA. We want to cast a wide net. You can talk about - we get newsletters out, we have 10s of thousands accessible like that, through the printed media that FEMA use. They have netted a large number of the nominations in this first compendium. We were worried by advertising on Internet that we would be swamped. What if panel could only handle a fraction. Compendium is a continuing compendium. There will be another one. It is a purposeful decision of ours not to publish and therefore not to review everything. This is not scientific. It is decisions made by people applying common sense to the issue of dealing with the reading public. When someone has written a great idea but not demonstrating empirically it s success, we write to the m and tell them to write s when they have the data. 80 to 90 of emergencies in the field are handled by local government.

AHCPR. Our guideline panels review published and fugitive literature. We also supplement with statistical experts and technical writers. Ancillary helpers in the process.

SAMHSA. first you convene an expert panel to determine how many you are going to do. Originally 10 highest procedures - costliest - under Medicare. Fugitive literature many times has better data for diffusion than published literature because it has negative results. No guidance on what you select. Suggests looking at history of AHCPR.

FEMA. A key question is who is the customer. In FEMA, customer is broad population, not congress, not executive branch. We were not looking for technical experts to judge the quality of the program.

Questioner: Carole LaCampagne, OERI: Topic focused issues. Carole coordinates the development of the Mathematics and Science Expert Panel.

One issue is the idea of promising vs exemplary. There are limitations when you do this, require a certain methodology, statistical, before and after, you limit yourself to small things. If I teach this one week unit and then measure afterwards - but if we are talking about broad changes in curriculum, these are very large and broad things. How do we account for things we don't have time to evaluate well but are worth getting into the public arena. Additionally, why are we doing this anyway? Our main purpose has to be a consumer guide but that consumer is not one type of person - it can be a teacher, a college professor, the legislature, Congress. They want to know what is the good stuff and they don't want to know all the good stuff. How do we make this manageable, identify the good but keep it within a manageable realm? We are talking about expert panel but what about a different problem - a journal - I submit an article to a journal and it is given to somebody on the editorial panel whose expertise is similar to mine - there are journals that appeal to me, idea of journal model, does that have applicability.

NSF. Lots of agencies are doing wonderful things in math, do we review their stuff, do we accept their standards, how do we link with other agencies, so we can collectively provide good information. Feedback. This is really needed. Is it meaningful, will it get out of hand, or could we have electronic feedback from teachers - that may be very valuable information for other users too. Appreciate a few answers now.

AHCPR. When was at the Cancer Institute with online treatment information system, we ended up with a monthly journal club. Our charge was to look at state of art and promising. We had 100 different cancer types - we put together an editorial board - initially all physicians and token nurse - each month we scoured the literature - reputable journals, key people publishing, look at stat things - randomized trials - group felt making decisions on promising anecdotal was not best way - also a section called investigational. Like a journal club. It worked. Continuous updating process.

SAMHSA. I have seen both models work jointly and separately. Journal model should be first step and then you need larger committee to come together. Model that works best is one just described that involves a continuous interaction among people.

AHCPR. We can talk later. We do have a grant doing something on feedback.

Questioner: Allen Schmieder, OERI. Issues related to the optimal uses of technology. We in sixties formed expert panels but included customers from the beginning. When NDN was formed, we decided net wasn't large enough and decided most good practice could not meet criteria - true now. We talked about sharing effective practice that consumers want without evaluating it. I would ask two basic questions: how will new technologies affect process of identifying, sharing exemplary practices? How do we get feedback? Second technology question is won't increased integration of technology affect the way you look at practice? How will way we look at process change because of integration of new technologies? My view is we are looking in wrong closets. One of most important contexts is whose closets are we looking in? Cutting edge of knowledge is where we ought to be looking in terms of knowledge to be shared.

SAMHSA. New technology will make it easier. I am concerned tremendously that information obtained through the Internet itself becomes a contaminant in the process. Then who is the consumer? Teachers, parents, state legislators, etc. How do we know we get the right information to them? We have about 17 different databases and WWW pages with almost no editorial content control. What technology will allow us to do is make information good and bad available to general public. We can't decide how technology will change process. We also have to be careful that technology is not in every school system. We can also use technology to disenfranchise. We have to be careful that to put in place something that can be used will disenfranchise those who don't have access to it.

Michael Sikes, National Endowment for the Arts - We have been trying to solve the wrong problem. The problem is not that we need more information, it is that we have too much information. What AHCPR said is we are confronting a mountain information. WWW - is a tool through which we can access vast amount of information. With WWW, we can more easily access information but we don't have basis for judging its quality. New world will ask disseminees to evaluate what they are getting.

H

01/19/1996

COMMISSIONED PAPER:

"STANDARDS AND THEIR USE IN THE FOOD AND DRUG ADMINISTRATION (FDA)"

GERALD L. BARKDOLL DPA

I. INTRODUCTION

The 1994 federal legislation reauthorizing the Office of Educational Research and Improvement (OERI) prescribed the development and use of standards for a variety of tasks ranging from the review of grants and contracts, to the identification of promising programs, products, practices and policies for recognition and support.

Standards have been used in the Department of Education (ED) for many years. The current initiative is designed to build on past successes and to overcome barriers that have limited the effective use of standards in ED initiatives. As part of this renewed interest in the use of standards, a number of initiatives have been undertaken to reflect past experience and to explore activities in other organizations that may provide valuable models and opportunities.

The U. S. Food and Drug Administration (FDA) has been identified as an agency that develops and applies a variety of standard setting activities to a diverse set of programmatic activities and responsibilities. The purpose of this paper is to describe several standard related activities in the FDA that may have relevance to OERI's renewed efforts.

II. OVERVIEW

FDA as a Resource

The FDA is a potentially promising focus for exploration for three reasons. The first is that the FDA has been "in business" for a long time, and has used standards for many of its activities. Second is the agency's high public visibility and its involvement in highly controversial issues of critical importance to the public. Third is the diverse and often contradictory forces that impact on agency decisions.

The legislative mandates assigned to the FDA have evolved for nearly 100 years. Although 1906 is often cited as the founding date for the agency, many of its initial activities had their beginnings in the U.S. Department of Agriculture. Over the years, new laws assigning new responsibilities to the agency have typically been precipitated by a highly visible failure in the market place. These market failures have placed consumers at risk of economic fraud and health-threatening products or practices. These events range from the life-threatening food processing practices common in the early 1900's, to more recent public concern over AIDS contamination of the nation's blood supply, and periodic drug tampering episodes.

The specificity of the agency's legislative mandates has changed dramatically over the years . Early laws, in essence, instructed industry to "do the right thing" and directed FDA to make sure that they did. Many of the agency's laws empower it to seek punitive actions including injunctions, seizures of contaminated products, and even fines and prison terms for offenders. Recent legislation has been much more prescriptive and focuses on the agency's processes. For example the 1976 amendments to the Food, Drug and Cosmetic Act prescribe how the agency was to regulate Medical Devices. These amendments included specific instructions as to how the agency should classify medical products by degree of risk, how products in each of these groups was to be regulated, the process for reviewing products already on the market, and the frequency of inspections of medical device manufacturing facilities, as well as other agency activities.

FDA and Ed decisions impact on, and are impacted on by many stakeholders. In FDA's case, consumers, the regulated industry and health professionals are three of the most active stakeholder communities. The agency's relationships with members of these communities are complex. For example:

- Consumer activist organizations support divergent and sometimes mutually exclusive goals. Some groups support increased regulation (and reduced or delayed marketing) of food additives, new prescription drugs, and medical devices, while other consumer organizations support rapid approval and increased availability of new products.

- Physicians and other health professionals share FDA's responsibilities for the safe and effective use of many products. Many drugs cannot be obtained by consumers without a prescription written by a doctor, dentist or veterinarian. In addition, many medical devices are used exclusively by physicians including some who must acquire special credentials, e.g. radiologists.

- The historic relationship between FDA and industry has been one of "regulator" and "regulated." That relationship has become more complex as contemporary legislative mandates have required the agency to be supportive of small medical device firms and have assigned the agency responsibility for funding the development of orphan drugs, i.e. drugs for rare diseases. The periodic anti-regulatory rhetoric of national political campaigns has also influenced the agency-industry relationship.

The role ambiguity precipitated by the conflicting, and constantly changing, expectations of FDA's stakeholders may be familiar to ED managers who must balance the desires of students, parents, teachers, educational administrators, the industry that develops and markets education materials, and others. ED managers may find themselves reconsidering the ED "mission," just as FDA managers have found themselves debating if the agency was a: (1) regulatory agency, (2) consumer protection agency, or (3) public health agency.

ED and FDA deal with issues having high public visibility and interest including the well-being of the nation's youth. For many years FDA managers have recognized the public's interest in the welfare of children. As far back as the early 70's the agency established planning processes that expressly identified three criteria for allocating agency resources. One of these criteria was titled "public sensitivity" and included the potential impact of agency activities on children and other vulnerable groups.

Standard Defined

OMB Circular No. A-119 defines a standard as "...a prescribed set of rules, conditions, or requirements concerned with the definition of terms; classification of components; delineation of components; specification of dimensions, materials, performance, design, or operations; measurement of quality and quantity in describing materials, products, systems, services, or practices; or descriptions of fit and measurement of size." International harmonization activities have addressed additional aspects of standards. For example, the General Agreement on Tariffs and Trade (GATT) attempts to reduce the unnecessary nontariff barriers to trade. On December 8, 1994, Public Law 103-465 was enacted in the United States to approve the Uruguay Round agreements. These agreements included the concept of volunteerism with respect to standards. Standard was defined as "[A] document approved by a recognized body, that provides for common and repeated use, rules, guidelines or characteristics for products or related processes and production methods, with which compliance is not necessary...." The term "technical regulation" was coined to define a similar document with which compliance is mandatory. This differentiation may be relevant to ED to the extent that standard setting is expected to be international in scope and impact.

III. THE IMPORTANCE OF STANDARDS IN FDA PROGRAMS

Agency Functions

The FDA is organized into Centers that correspond to its product responsibilities. Approximately 60% of the agency's resources are assigned to product centers (formerly bureaus) focusing on Foods, Human Drugs, Veterinary Drugs, Biological Products (including blood), and Medical Devices (including radiological emitting products). The remainder of the agency's resources are assigned to the agency's field force that conducts inspections and tests products across the country.

FDA responsibilities can generally be divided into two major functions: pre-market review/approval of new products, and post-market surveillance of products in use. Pre-market approval frequently includes a concurrent review of the labeling proposed for the product. These reviews are to assure that the labeling will provide the information necessary for the product to be used for its intended purpose by health professionals and/or by consumers. Although there is a great deal of coordination between the product specific centers and the field force, the majority of the pre-market approval is accomplished by the centers, and the majority of the post-market surveillance work is done by the field.

Cross Cutting Issues Involving Standards

A detailed and comprehensive review of the agency's use of standards could be accomplished by focusing in sequence on each of the agency's functions. A review of this type would identify substantial variations due to differences in laws and technology, as well as differences in the regulatory history of industries. For example, the food industry has regulatory experience dating to the beginning of the century while medical devices are still in the early stages of their regulatory experiences. The agency's long-term utilization of standards is demonstrated by the fact that Food Standards were an early approach to food regulation while the 1976 Medical Device Amendments engaged the agency in regulating a large genre of medical devices through the promulgation of standards.

A comprehensive review of the use of standards by the FDA is well beyond the scope of this paper, and probably well beyond the interest of any likely reader. Consequently, three cross-cutting topics have been selected on their basis of potential usefulness to ED. These topics involve issues that transcend the FDA products and functions. They are:

1. Process dimensions of product approval process. This topic encompasses the roles and relationships of the participants in FDA's review processes, the forces that shape these roles and relationships, and how these roles and relationships compare with those found in other countries.

2. The conflict between achieving benefits and managing risks. Balancing benefit and risk based on results has been an FDA responsibility for many years. This topic addresses some of the important issues that have emerged during that time.

3. The use of outside experts in the development and application of standards. Standard setting organizations and individual experts have played an important role in agency activities for many years.

IV. PROCESS DIMENSIONS OF PRODUCT APPROVAL PROCESSES

Human Drugs, The Principal Focus

FDA regulates products accounting for one-fourth of consumer expenditures. The average consumer comes in contact with dozens of FDA-regulated products during the course of the day. The size and diversity of the FDA responsibility threaten to unduly confound the discussion of standards, the principle topic of this paper. For that reason, one product area, human drugs, has been chosen as the principal focus of this paper. Selecting human drugs as the principal focus has several important advantages, including: (1) high public interest; (2) a large and diverse set of products accounting for over fifty billion dollars in annual sales; and (3) multiple processes involving government review of research findings. Other product areas (foods, veterinary drugs, medical devices) will be cited when they present a unique concept or relevant experience.

Reviewing research conducted by others is one of FDA's principal functions. Each year FDA scientists, physicians, regulatory experts, statisticians, and other professionals review tens of thousands of pages, tables, and charts contained in industry applications. These applications describe the results of studies designed to demonstrate the safety and efficacy of a wide variety of products and processes. As an example of the extent of data received by the agency, the Center for Drug Evaluation and Research received: (1) between 223 and 328 New Drug Applications in the years 1983 to 1993; (2) between 1337 and 2576 Investigational New Drug Applications (INDs) between 1983 and 1993; and (3) over 1600 changes and supplements to INDs each year from 1983 to 1994. The information contained in some of these applications fills dozens of volumes and frequently occupies over 10 feet of shelf space. The quantity of data contained in these studies has inspired some to suggest that FDA stands for Federal Data Administration.[1]

Economic Impact

Economics plays a critical part in FDA's review of research submitted by industry. On the one hand, FDA's review role does not encompass the intended selling price of the product. On the other hand, the implications of approval or rejection are of critical importance with stock prices, corporate executive compensation, and even the survival of individual firms riding on the agency's decision. A delay of a few months can easily cost a company several million dollars in lost profits from a new product.[2]

Corporations employ a wide variety of techniques to assure prompt review approval ranging from extensive training for their employees involved in the preparation of applications, to requests for their congressional representatives to intercede on their behalf. The agency's response to speed up the process is confounded by the imbalance of responsibilities and resources, the public outrage surrounding a dangerous product allowed on the market, and the need to assure that trade secrets are protected.

The economic implications for an individual firm can be immense, including the survival of start-up companies that are frequently undercapitalized. Congressional mandates require the agency to protect information pertaining to new products. For example, the fact that a New Drug Application (NDA) has been submitted to the agency for review is considered a trade secret. In some instances firms advertise the fact that they have a new product under review. In these cases the agency is permitted to confirm the fact that an NDA has been submitted. Protecting intellectual assets may be one of the issues that ED will have to face as it evaluates and supports the development of new programs and techniques. Assuring that the worth of the new programs are not threatened by the review process may be critical to a dynamic development environment.

Lessons from the European Union

Interesting insights can be gained by comparing U.S. practices with those of foreign countries. The European Union (EU) provides an interesting comparison since it includes 15 developed countries established by treaty to create a common market and promote political integration. The processes underlying the drug approval processes in the EU differ in informative ways from the US processes. In general, the EU approach tends to be more collegial and consultative, while the U.S. approach is frequently described as "arm's length." A recent study listed three important differences:[3]

- In Europe, regulators often come from industry, or move back and forth between industry and government, a practice discouraged in the United States.
- EU regulators require drug sponsors to prepare and submit Expert Opinions (assessments of the merits and shortcomings of their drugs) rather than raw data and detailed summaries. In the United States, it would be a conflict of interest for drug sponsors to assess their own drugs.
- European reviewers of Marketing Authorization Applications tend to use a "top-down" approach, basing their assessments on summary information and Expert Opinions. FDA reviewers, in contrast, tend to use a "bottom-up" approach, analyzing raw data or detailed data summaries to reveal their own, independent conclusions about the merits and shortcomings of the drug.

Some Derivative Questions For ED

The differences between U.S. and EU considerations of research findings are demonstrated by significant differences in the form and content of applications. Translating these differences to the research to be reviewed by ED suggests a number of questions.

1. New drug applications in the U.S. contain both nonclinical and clinical data compared to EU submittals which tend to emphasize clinical data only. To what extent will research reviewed by ED focus on the theoretical underpinnings of the research versus the empirical evidence?

2. New drug applications in the U.S. typically include more raw, pre-analyzed data, thus permitting FDA reviewers to perform their own analyses. To what extent will ED reviewers and researchers want to conduct similar analyses?

3. Instructions for proper use, typically referred to as "product labeling," is a critical part of U.S. submissions. EU submittals are less comprehensive and detailed. To what extent will ED review focus on the "instructions for proper use"?

4. Both U.S. and EU new drug applications include a description of the analytical methods used for product testing, and a description of how the product will be packaged. Are there similar testing and packaging issues ED will have to address?

V. THE CONFLICT BETWEEN ACHIEVING BENEFITS AND MANAGING RISKS

Safety First

The initial concerns of industry and FDA focus on the safety of a product under consideration. Only after the potential adverse reactions of a product have been identified and calibrated, with respect to dose, will testing proceed to determining efficacy.

Prior to 1962, FDA's legislative mandate extended only to the safety of human drugs. More than 3500 drug products were on the market at the time the efficacy requirement was passed. A special process called the Drug Efficacy Study Implementation (DESI) was established to review the efficacy of these marketed drugs. The process was costly and required many years to complete. Most observers would agree that it is more efficient to deny market access to ineffective drugs than it is to remove ineffective products once they have established a market niche and are making an important economic contribution to the sponsoring organizations.

Many of the products on the market in 1962 are gone. Dr. J. R. Crout, former director of FDA's Bureau of Drugs, made an interesting observation about the drugs that have not survived the DESI process. He commented, "The review successfully wiped out all irrational combinations and, indeed, more combinations with more than two ingredients.... It was to bring honest labeling and scientific integrity to all the drugs that passed the review."[4]

Fraudulent or bogus products present problems closely akin to safety. They also engender a very high "outrage factor" in the public, since they are usually associated with charlatan promoters ripping off the public. Unsafe or ineffective products that reach the marketplace can, in fact, have a direct impact on the public's safety since they: (1) delay the start of truly effective treatment; (2) divert scarce public resources away from effective treatment; and (3) divert agency resources from legitimate new product review and approval to damage control activities.

Determining Efficacy

The 1962 amendments required that drug effectiveness was to be demonstrated by adequate and well-controlled trials. Those high-sounding words were subject to many different interpretations, and consequently produced many tests of will between industry and the agency. After a number of years of case-by-case decision making, the FDA published regulations describing the operational dimensions of adequate and well controlled. These included the need for a control group, randomization, blinding, and proper statistical analysis. Finally, by 1989, Dr. Crout was willing to observe, "Today the revolution is complete. The drug industry conducts only controlled clinical studies in support of effectiveness. A scientific community of physicians and statisticians specifically concerned with trial design and analysis has blossomed. There is now a Society of Clinical Trials.... Medicine is at last acquiring solid information on the effectiveness of its remedies."[5]

Over the years, FDA reviewing officials have become more comfortable providing guidance to industry researchers. Industry representatives have constantly asked agency officials to establish the standards for acceptable research, and from their perspective additional guidance has no doubt been long overdue. Two explanations for the apparent slow development of guidance are: (1) the science of drug testing has continued to evolve and (2) the agency's resources to review applications have always been exceeded by the workload on hand, thus discouraging any diversion of resources to activities not mandated by law.

One of the more successful initiatives to provide guidance to industry researchers has been "Points to Consider" developed by FDA's Center for Biological Research and

Evaluation (CBER). Each of these documents is highly specific to products and processes. For example, on August 22, 1995, CBER published a 15-page "Points to Consider in the Manufacture and Testing of Therapeutic Products for Human Use Derived From Transgenic Animals." The letter transmitting this document to industry stated, "These 'points' are not regulations nor are they guidelines [both of which have well defined legal meanings and force], but represent the current thinking that the Center for Biological Research and Evaluation staff believe should be considered at this time." Among other topics the document addressed information to be provided to the agency including: (1) detailed characterization of the original gene to be introduced into the animals; (2) the method used to introduce recombinant DNA into animals; (3) the methods used to determine that the founder animal is producing the desired product within acceptable criteria; (4) feeding of animals used in production; and (5) actions to assure that no endogenous or adventitious agents contaminate the process.[6]

The standards provided in the points to consider documents, like most standards, are subject to constant revision as scientific knowledge continues to expand at an ever-increasing rate. For example, a "Points to Consider Document on the Characterization of Cell Lines Used to Produce Biologicals" was published July 12, 1993, to replace a document of the same title issued in 1987. In addition, the document stated:

> Advances in biotechnology are occurring rapidly ... therefore, information in this document is subject to change as new and significant findings become available.... Therefore the Center for Biologics Evaluation and research will review the adequacy of testing of any cell line on a case-by-case basis."[7]

Balancing Risks and Benefits

The potential harm associated with pharmaceutical products has necessitated a carefully sequenced exploration of risks and benefits. For example, all practical forms of computer simulation and animal tests are undertaken before an experimental drug is tested in humans. Although the use of computer simulation continues to grow, laboratory animals are needed to determine what the drug does to the body, and what the body does to the drug. For example the amount of drug actually absorbed into the body may vary a great deal from animal to animal and subsequently from human to human. Experienced classroom teachers would no doubt observe that the amount of information absorbed also varies a great deal from human to human.

Human testing typically involves three phases. Phase 1 usually involves healthy volunteers and attempts to determine what happens to the drug when it enters the human body--how it is absorbed, metabolized and excreted. Phase 2 testing involves up to several hundred patients and typically lasts from several months up to 2 years. Phase 2 testing helps confirm short-term safety, but is primarily aimed at determining efficacy. Only about 45% of the drugs that originally entered Phase 1 testing survive this far. Finally, Phase 3 testing involving hundreds or thousands of patients is undertaken. The focus is on efficacy and the testing attempts to approximate normal medical practice to the extent possible. (8)

FDA's responsibility is to determine if new drugs are safe and effective. The legislative mandate does not require the agency to determine if a particular product is the safest, or the most efficacious. Neither does it require that the agency determine that a proposed new product has the best benefit/risk ratio of the products available to the public. It is left to the market place to sort out the "best", consequently companies invest a great deal in sales and marketing efforts. FDA 's legislative mandate absolves it from dealing from two issues that may be significant to ED:

1. Will the standard for measuring the worth of new educational programs be relative or absolute? That is, will new programs be compared with an independently established benchmark, or against programs now on the market?

2. How will ED identify new programs it considers effective? For example, will it issue the equivalent of a "Good Housekeeping Seal of Approval" to products it considers adequate?

Given the potential economic consequences of new product approvals, a few individuals are tempted to omit evidence of safety problems, or artificially augment evidence of efficacy. These falsifications are frequently discovered by agency reviewers when the data looks "too good to be true." The penalties for submitting falsified official documents to the government are severe, and include disbarment of clinical investigators as well as fines and imprisonment for corporate officials. The evidence submitted by the firms serves as the basis for the approval decision, and is the foundation for the labeling that communicates the product's benefits and risks to health professionals and the public. FDA not only approves the exact labeling to be used on the product, but closely monitors advertising and educational programs sponsored by industry.

The Conflict Between Good Answers and Quick Answers

One of medicine's most celebrated clinical trials was conducted by Louis Pasteur. He treated patients who had been exposed to rabies with an experimental anti-rabies vaccine. All of the treated patients survived. This contrasted with the 100 percent fatality rate that was known to be associated with rabies. Drawing a conclusions was easy.

Unfortunately, most clinical trials are not so straightforward. Drug testing is complicated by the nature of many illnesses. Minor injuries, colds, and some medical problems can be expected to run their course in a predictable time frame. Unfortunately, chronic conditions like arthritis, multiple sclerosis, and depression are much less predictable.

Determining the effect of a particular treatment on these types of illnesses is much more difficult.

Some drug reactions (both expected and unexpected) appear immediately; however there is always the concern that important long-term effects will emerge and change the risk/benefit ratio. The uncertainty surrounding the ultimate outcome and the time required to conduct "well controlled" research are in direct opposition to the public's interest in receiving prompt and effective treatment.

The scientific and regulatory communities have developed a number of strategies for ameliorating the conflict between speed and accuracy in pre marketing approval processes. These include surrogate end-points, Phase IV trials, treatment INDs, and the collection and analysis of adverse drug reactions. Each of these strategies may suggest an approach for ED to use in balancing the identification of new educational techniques that are truly effective, versus the need to promptly find and promote effective new techniques.

- Surrogate End-Points. Surrogate end-points are laboratory indications or a physical sign that indicates the drug is having a desired effect, although the patient may not feel or function better. High blood pressure and elevated serum cholesterol have been linked to heart and blood vessel diseases. Drugs that can be demonstrated to have an impact on blood pressure and the level of serum cholesterol may be approved for marketing even though direct evidence of the drug's effect on angina, congestive heart failure, and paralysis following a stroke have not yet been demonstrated. As a condition of approval, FDA may limit the distribution and use of the new product to settings where additional confirmatory evidence can be collected.

- Phase IV Clinical Trials. The U.S. and European approaches provide a contrasting perspective on decision making before all the important risk/benefit data is available. In instances of life-threatening or severely debilitating illnesses, FDA may allow the drug to enter the market, but on the condition that the sponsors conduct Phase IV studies. These studies focus on the longer term as well as infrequent but serious adverse reactions. In comparison, regulatory agencies in the European Union can approve a new drug for a finite period of time, with renewal dependent on the sponsor's successful completion of additional trials. In the U.S. the burden of proof to support a drug withdrawal from the marketplace falls on the FDA. In Europe the burden of proof for renewal falls on the sponsor. To what extent should ED assure that ineffective, or perhaps even harmful educational approaches be similarly removed from the marketplace?

- Treatment INDs. Treatment INDs are a variation on the formal drug testing process described earlier in this paper. The rapid spread of AIDS and the urgency to treat this deadly illness provided the impetus for aggressively exploring alternative approaches to drug testing. The earliest experience involved zidovudine (AZT). A summary of this experience was described in the October 1993 issue of the *FDA Consumer.*

> Initial (Phase I) testing of the drug in 33 patients with AIDS carried out between July and December of 1985, yielded encouraging results. Phase 2 trial to assess the drug's safety and effectiveness began in February 1986. About 300 people with AIDS at several centers around the country were randomly selected to receive either AZT or a placebo.
>
> These studies were abruptly halted in September 1986 when it was discovered that 16 patients receiving the placebo had died, while only one death had occurred among those receiving AZT.... As a result more than 4,000 AIDS patients were treated with AZT before its approval as the first anti-AIDS drug under the brand name Retrovir in March 1987.[9]

The use of the IND process can jeopardize the ongoing research needed to assure the efficacy of a new product. For that reason the FDA issued regulations in May 1987 that spell out the conditions for investigational IND designation. The regulations require informed patient consent and the continuation of carefully controlled trials, and prohibit the commercialization of the product.

Monitoring adverse reactions. The FDA receives over 50,000 reports of adverse drug reactions per year. The majority of these reports come from drug manufacturers who are required by law to submit reports. Additional information is obtained through a number of contracts with Medicaid and health care organizations. This database is aggressively scanned for new trends with special emphasis on drugs recently introduced onto the market. The system is credited with the prompt identification of adverse drug reactions including the suprofen flank pain syndrome, as well as excessive sedation and respiratory depression associated with midazolam. The system has been credited with identifying excessive adverse reactions not identified in other countries where the product is marketed.[10]

Diverse Populations

Research to determine the safety and efficacy of new products depends in part on the homogeneity of the test population as well as the population of intended use. Determining the risks and benefits of a product in segments or subsets of the population greatly complicates the testing process. Demonstrating that a particular product or intervention is effective in a population of white adult males does not necessarily prove safety or efficacy in children, women of child bearing age, and other groups with unique characteristics. The inclusion of women in clinical trials is an example. In 1993 FDA published guidelines encouraging drug companies to include adequate numbers of women in clinical trials and to pay particular attention to factors such as phases of the menstrual cycle, menopause, and use of oral contraceptives. These guidelines operationalize a reversal of earlier agency policy that prohibited testing of drugs in women of child bearing age.[11]

It seems likely that ED will find itself struggling with one-size-fits-all solutions versus solutions tailored to meet a variety of student needs and public expectations. A single solution will certainly be more attractive in the political arena, and may be more effective in attracting needed resources. Unfortunately, those who control the resource allocation process may be unwilling or unable to relate to the diversity of cultures that the educational experience must accommodate. This could easily be the most difficult issues ED will have to address.

VI. ENGAGING EXPERTS IN THE ESTABLISHMENT AND APPLICATION OF STANDARDS

FDA recruits health professionals, attorneys, scientists from many fields, and public administrators to meet the mandates prescribed by Congress and perform a public service role expected by the public. The diversity of the agency's responsibilities make it impossible to employ all of the necessary specialized personnel. As a matter of necessity (and occasional congressional mandate), agency activities engage established standard setting organizations as well as individual outside experts in the establishment of standards, and in the application of these standards to the agency's work.

Required Industry Testing

The many experts and substantial resources found in the regulated industry make it a likely place for the agency to turn in its attempts to assure safe and effective products. For many years the agency has established and published the standards industry is expected to meet, and has required the industry to do the necessary testing. In a number of cases the establishment of the standard extends beyond the initial decision that a product can be marketed to the ongoing testing of the product. For example, some biological products are tested on a lot-by-lot basis prior to release to the marketplace. The relevant federal regulations read:

> No lot of any licensed product shall be released by the manufacturer prior to the completion of tests for conformity with standards applicable to each product....

The degree of control available to the FDA is substantial since the regulation goes on to state:

> Samples of any lot of any licensed product ... together with the protocols showing results of applicable tests, may at any time be required to be sent to the Director, Center for Biologics Evaluation and Research. Upon notification ... a manufacturer shall not distribute a lot of a product until the lot is released by the Director....[12]

The regulation goes on to describe the test process in great detail, including the: (1) test animals (overtly healthy guinea pigs weighing less than 400 grams each and mice weighing less than 22 grams each); (2) test procedures (the test shall go on for 7 days for both species); and (3) test requirements (the test animals survive, and weigh no less at the end of the test than at the beginning).

Standard Setting Organizations

The value of generally recognized and accepted standards is demonstrated by the existence of many organizations devoted to this activity. Some of the better known organizations include: American National Standards Institute, American Society of Testing and Materials, Association for Advancement of Medical Instrumentation, and National Committee for Clinical Laboratory Standards and Underwriters Laboratory. The hundreds of standards established by these and other organizations range from standards for Dental Casting Gold Alloy and Cardiovascular Catheters to Standard Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture. These and similar organizations are funded primarily or exclusively through the sale of standards and related activites, e.g. the testing of products against standards.

The historical significance of the use of outside experts can be demonstrated by a brief review of the Technical Electronic Products Radiation Safety Standards Committee (TEPRSSC). The TEPRSSC was established by Congress in 1968 to provide consultation before the FDA Commissioner prescribed any performance standard for electronic product. The committee's functions were specifically listed as:

1. May propose electronic product radiation safety standards to the Commissioner for consideration;

2. Provides consultation to the Commissioner on all performance standards proposed for consideration under 42 U.S.C. 263 f; and

3. May make recommendations to the Commissioner on any other matters deemed necessary or appropriate in fulfilling the purposes of the act.[13]

TEPRSSC consists of 15 members appointed by the Commissioner after consultation with public and private organizations with an interest in the safety of electronic product radiation safety. The law is prescriptive with respect to the stakeholders represented. The committee must include five members from government agencies (including State and Federal), five members from the regulated industry, and five from the general public including one representing organized labor.

Assuring Credibility

The objectivity and neutrality of standard setting organizations is paramount if the standards they set are to be used and trusted. Many FDA employees are experts in their chosen fields and consequently are often called on to serve in official and liaison capacities in standard setting organizations. The agency has recognized the importance of assuring that its employees' reputations are above question and has established strict guidelines for employee participation. These guidelines serve as valuable criteria for any organization considering the use of independent organizations in standard setting activities.

An FDA employee may be permitted to engage in the activities of private standard setting organizations if the organization requests the participation and clearly articulates the role of the organization and the expected role of the agency employee. In addition, rules will apply to the relationship:

1. The organizations must explicitly state that the employee's participation does not connote official agency endorsement of or agreement with any decisions made by the organization.

2. Participation by the agency employee precludes subsequent service as the deciding official on the standard if it should later come before the agency.

3. The request for this and all other correspondence involving the employee's participation will be made part of a public record available to the public.

4. All organizational activity involving the FDA employee will be based on sound scientific and technological information, will permit revision on the basis of new information, and will be designed to protect the public against unsafe, ineffective or deceptive products or practices.

5. The activity of the organization will not be designed for the economic benefit of any company, group, or organization.

6. The group or organization must have a procedure for accepting and considering information from the public, without payment of fees by the public.[14]

Institutional Review Boards (IRB)

One of the primary concerns of conducting research on humans is that the human subjects fully understand the risk to which they will be exposed. This concern, as well as other characteristics of high quality research, are determined principally in the early design stages. Quality cannot be retrofitted after the research has been completed.

The 1962 amendments required that patients be informed. The FDA subsequently published regulations prescribing the early and continued involvement of IRBs. One informed observer believes that the system is working well. He states:

> "These [Institutional Review Boards (IRB)] include as members not only physicians but also lawyers, churchmen and other members of the non-medical public. Members of the IRBs have an outstanding record of taking their work seriously and I believe they have contributed enormously to maintaining the credibility of the clinical research process. Without the informed consent and the IRBs, there is little doubt that drug research in the US would be conducted under a burden of suspicion, continued press inquiry, and litigation. The IRB system also has improved the quality of review brought to protocols, and the sensitivities of clinical investigators and the drug industry in providing full information to patients in clinical studies. We are all indebted to the contributions of these groups in the clinical research process."[15]

Advisory Committees

FDA has established 17 committees of experts to advise it on the safety and effectiveness of drugs. Additional committees have been established to advise the agency on foods, veterinary drugs, biological products and medical devices. The outside experts add a wider spectrum of judgment and experience to tough issues facing the agency.

Advisory committees include health professionals with particular knowledge. For example, the FDA's Fertility and Maternal Health Drugs Advisory Committee includes seven physicians specializing in obstetrics, gynecology, reproductive biology and epidemiology. The nursing profession and behavioral scientists are also represented. This committee, like all FDA advisory committees, includes members nominated by consumer organizations.

Committees meet two to four times a year depending on the level of activity in their area of expertise. Committees frequently advise the agency on the information to be included in new product labeling and are often called on to provide an opinion on the adequacy of information submitted by a new product sponsor. An example of the valuable work of a committee involved Cognex, a drug designed to treat Alzheimer's disease. The committee concluded that initial clinical trials did not demonstrate product effectiveness. The committee recommended the drug be tested at a higher dose and recommended that the agency allow the drug wider distribution as a treatment IND. The drug was subsequently made available to over 7,400 patients and was later determined to provide a small, but meaningful benefit for some patients with mild-to-moderate Alzheimer's disease.

Advisory committees are frequently called on to review adverse reactions reported for a previously approved drug. As a result of such a review the committee may recommend changes in dosage, increased label warnings, or denial of the drug to certain vulnerable patient populations.

Experts and Conflicts of Interest

Bringing experts in from the "outside" involves many of the same issues as the involvement of FDA employees in outside standard setting activities. Many of the experts the agency would like to call on have developed their expertise through affiliations with drug firms and universities conducting industry-supported research. Committee members are subject to the federal laws that prohibit participation in an official action in which they have a financial interest. For example, if a committee member is on the faculty of a university that has a grant from a firm, the committee member cannot act on issues pertaining to any of the firm's drugs. Before each meeting, committee members are contacted to determine if they own stock or have grants or contracts involving issues or firms on the agency. Some waivers are allowed if the stock holdings or grants and contracts are small, and the need for the committee member's expertise is essential.

The "Government in Sunshine Act" of 1977 dictated that advisory committee meetings would be open to the public except when confidential, commercial, or trade secret information, or law enforcement investigations are discussed. Many agency managers and committee members were concerned about the dilatory effect that open meetings might have on scientific discussions. They were current in presuming that the meetings would be attended by the public, although they could not have predicted the make-up of the audiences. The most visible public observers are stock analysts who continuously leave the room to report the latest good or bad news about any product or company under discussion. Stock prices of drug companies often fluctuate markedly based on the reporter's latest message to their firms.

All in all, the advisory committee process appears to be working well. In the opinion of one participant who has seen the process from outside as well as inside the agency:

> It is now clear that those fears were unfounded. Although there have been examples of meetings that were poorly run, of participants who were less than forthright and of press excesses, these problems have proven in

> general to be anomalies. The great majority of FDA advisory committee meetings proceed as well organized and thoughtful discussions by serious people. As a citizen and an employee of the drug industry today, I am pleased by the openness with which the FDA advisory committee system operates. Much can be learned from these meetings--the rules, kinds of trials the agency wants, the kinds of statistical analysis that are persuasive, the kinds of evidence needed to support certain claims. All can profit from the mistakes of others, so that future meetings in front of the committee can be approached with the confidence that, if one's homework is thorough and the issues are fairly addressed, arguments will be listened to and understood.[16]

VII. A POLICY ON STANDARDS

On October 11, 1995, the FDA published a Federal Register notice of its Policy on Standards as it related to international harmonization activities.[17] The policy statement was based on a draft policy published a year earlier and the comments received from ten organizations including industry, standard setting organizations, consumer groups and trade and professional organizations. The final policy statement has a number of characteristics that may be of particular interest to ED as it considers its own standard setting activities. The policy:

1. is very clear about the purpose of standard setting, i.e. to assist the agency in implementing the law with respect to safeguarding the public health.
2. lays out certain conditions under which standards and standard setting will be considered appropriate, e.g.
    a. the standard is based on sound scientific information,
    b. the development process for the standard is open to public scrutiny,
    c. the standard setting process is consistent with the codes of ethics that must be followed by FDA employees.
3. asserts that the agency is not necessarily bound to use a standard developed with FDA participation. For example, the agency will not use a standard when doing so will compromise the public health.
4. identifies a variety of uses for standards including:
    a. inclusion in guidance documents for clinical and non clinical trials,

b. conducting the reviews of applications submitted to the agency,

c. conducting reviews of research protocols proposed by industry, and

d. inclusion of the standards in memorandum of understanding with other government agencies.

END NOTES

1. Center for Drug Evaluation and Research. "Offices of Drug Evaluation Statistical Report." Document Management and Reporting Branch, Office of Management, Food and Drug Administration, 1993.

2. Barkdoll, Gerald, and Kathleen McEvoy. "Predicting the Payoff from Earlier Drug Approvals." Pharmaceutical Executive. 7(11), November 1987.

3. Eastern Research Group, Inc. Developments in International Pharmaceutical Regulation: Implications for the United States (Final Report (p. 3-2) (Contract No. 223-91-8100). Prepared for the Food and Drug Administration, September 12, 1995.

4. Crout, J. R. "Objectives and Achievements of Regulations in the USA." In S. R. Walker and J. P. Griffin (eds.), International Medicines Regulations: A Forward Look to 1992 (p. 119). Lancaster, UK: Kluwer Academic Publishers, 1989.

5. Court, J. R. "Objectives and Achievements of Regulations in the USA." In S. R. Walker and J. P. Griffin (eds.), International Medicines Regulations: A Forward Look to 1992 (p. 120). Lancaster UK: Kluwer Academic Publishers, 1989.

6. Center for Biologics Evaluation and Research. "Points to Consider in the Manufacture and Testing of Therapeutic Products for Human Use Derived from Transgenic Animals" [Docket No. 95D-0131]. Rockville, MD: Food and Drug Administration (August 22, 1995).

7. Center for Biologics Evaluation and Research. "Draft of Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals [Docket No. 84N]. Rockville, MD: Food and Drug Administration (July 12, 1993).

8. Fleiger, Ken. "Testing in Real People." FDA Consumer, November 1987, 21(9): 11-13.

9. Fleiger, Ken. "FDA Finds New Ways to Speed Treatment to Patients." FDA Consumer, October 1993, 27(8): 16.

   Temple, R. "Objectives and Achievements of Regulations in the USA." In S. R. Walker and J. P. Griffin (eds.), International Medicines Regulations: A Forward Look to 1992 (p. 131). Lancaster, UK: Kluwer Academic Publishers, 1989.

   Fleiger, Ken. "Testing Drugs in People." FDA Consumer, August 1994, 28(6): 16-19.

12. Code of Federal Regulations 21 CFR 610.

13. 21 CFR 14.120.

14. 21 CFR 10.95.

15. Crout, J. R. "Objectives and Achievements of Regulations in the USA." In S. R. Walker and J. P. Griffin (eds.), International Medicines Regulations: A Forward Look to 1992 (p. 121). Lancaster, UK: Kluwer Academic Publishers, 1989.

16. Crout, J. R. "Objectives and Achievements of Regulations in the USA." In S. R. Walker and J. P. Griffin (eds.), International Medicines Regulations: A Forward Look to 1992 (p. 122). Lancaster, UK: Kluwer Academic Publishers, 1989.

17. Food and Drug Administration. "International Harmonization; Policy on Standards Notice." Federal Register Notice, October 11, 1995, 60(196).

18. "Draft Policy Statement on Industry-Supported Scientific and Educational Activities." Federal Register Notice, November 27, 1992, 157(229).

I

Review of Foundation, Associations and Non-profits Practices

In Designating Promising and Exemplary Programs

by Janet Carter and Diane Schilder

Section One: Background and Organization

The Office of Educational Research and Improvement (OERI) in the US Department of Education (ED) commissioned this paper to document the efforts of foundations, associations and non-profits in designating promising and exemplary programs. The paper focuses on selected recent and on-going non-governmental efforts which seek to identify what works in and outside of education. We ask what might be instructive for OERI's emerging System of Expert Panels from these experiences. As an internal OERI paper explains, "The System of Expert Panels is based on the assumption that many worthwhile products, programs and practices have already been developed ." [1]

We discuss how the strengths of such efforts can be replicated and describe the political, technical and procedural limitations in applying foundation efforts to the public sector. This paper is organized in five sections. The first is this overview. The second section describes the data collection, analysis and reporting strategy. The third section describes the political, technical and procedural limitations in applying principles used by foundations, associations and non -profits to the public sector. The fourth section provides an overview of approaches to determine effectiveness, and finally, we discuss the suggestions that may be applicable to the public sector.

An appendix contains information on selected projects.

Section Two: Data Collection, Analysis and Reporting Strategy

The data collection and analysis strategy began with key informant interviews of program officers at national foundations, directors of national associations and key informants at non -profit agencies. The analysis relies primarily upon these sources. The authors conducted searches of the ERIC database as well as databases of social and behavioral sciences. The review of the literature produced few articles in academic journals that are specifically described as foundation-funded efforts in developing criteria for determining the effectiveness of innovations. Further, foundation , association and non -profit reports and unpublished manuscripts are often not included in academic databases.

Key informants were identified from a list of participants at the Creating Change: Program Replication and Transfer conference arranged by the Northern California Grantmakers Association and held in Washington, DC in September 1993. The list of participants included

foundation staff, academics, non profits, and some government representatives. In addition, we conducted key informant interviews with members of national grant-making associations, such as the Grantmakers Evaluation Network and with experts who were identified by interviewees as key informants in the areas of replication of education and related initiatives.

The key informants we interviewed recommended a number of foundations, associations and non-profits. Based on these recommendations, we interviewed over 25 contacts representing these efforts and obtained documentation about their evaluation strategies, methods for collecting data for replication purposes, and dissemination activities. After these interviews, we narrowed the selection to 14 efforts that were involved in the systematic setting of standards, collection of data and dissemination of information (See matrix).

---

[1] Klein, Susan, "A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop and Share Promising and Exemplary Products and Programs" DRAFT 2/12/97.

Section Three: Political, Technical, and Procedural Limitations in Applying Principles Used by Foundations, Associations and Non-profits to the Public Sector

The process of making decisions about program "effectiveness" (or promising and exemplary programs) and disseminating such information to the public at foundations, associations, and non-profits differs from public funders. Specifically, foundations differ from public funders in three important ways. First, foundations perceive a need of privacy from the public concerning entities who are not designated. As one foundation staff person explained , "It would embarrass those who were not chosen" if the process of selection was open to scrutiny. Second, foundation staff may be selective in how they characterize the total review process. While expert panel findings, for example, are fully discussed with Board members, the overall review process "is more subjective" then is commonly acknowledged to these decision makers. Third, unlike publicly funded programs that tend to be prescriptive and are likely to have a single source of funding, initiatives that are foundation funded often receive support from multiple foundations, making it very difficult to "disentangle" the effectiveness of one foundation's contribution.

Associations and non-profits also differ from public sector agencies in important ways. Like foundations, some associations and non-profits are sensitive to private issues such as the research demands requested of local chapters in a membership organization. In addition, these organizations often have multiple sources of funding and each funding source may have different evaluation requirements. Thus, decisions about "effectiveness" within associations and non-profits may vary depending upon the foundation funding the specific effort.

Despite the limitations in applying principles drawn from foundations, associations, and non-profits to the public sector, some important lessons can be shared. Foundations, associations, and non-profits use a variety of approaches in determining promising and exemplary programs including using innovative evaluation methodologies to obtain quick turn-around data for decision-making purposes. Lessons learned are described in the section below.

Section Four: Description of Foundation, Association, and Non-Profit Methods of Determining Effectiveness

Foundations, associations and non-profits use a variety of methods to determine exemplary and promising initiatives and set different standards for making these decisions. The methods used to determine "best practices"; to make decisions about new areas for funding; and to evaluate initiatives they have funded differ across and within organizations. These different approaches reflect the values and missions of the organizations.

They frequently set criteria for judging the quality of different efforts based on their internal vision and mission and will then judge efforts against these criteria. Collected through expert panels, qualitative case study designs, participatory approaches, key informant interviews, and information that is informally collected, such data are used as sources of evidence for making funding decisions and for highlighting initiatives for others to replicate. For example, one foundation reported that its board determined the general direction for future funding and for determining whether initiatives were promising and exemplary to them. The staff translated the board's mission into a set of guiding principles. Programs applying for future funding were judged against this set of guiding principles. Programs that had already been funded were asked to report data that could be checked against the set of principles. Such data were not simply quantitative evaluation data. Included were descriptive information about the program's goals and accomplishments. In this instance, the foundation funded external evaluations only for large initiatives that were designed as research and demonstration projects.

In addition to using different methodologies for assessing promising or exemplary initiatives, the efforts included in this review have developed a number of approaches for disseminating information to consumers about "best practices". These include developing expert panels to determine standards for future foundation funding, setting u p award systems whereby standards are publicly stated and applied, and awarding recognition to innovative practices. For example, the Carnegie Council on Adolescent Development fostered by the Carnegie Corportation was developed to review the accumulated knowledge in the field and then to refine and state standards to which proposed projects would be "mapped on". The Innovations Award program carried out by the Kennedy School at Harvard, created by the Ford Foundation, is an example of an effort to develop standards for public agencies by holding up outstanding examples from state and local , and more recently, federal , projects and practices. The Dana Foundation has similarly developed a process for awarding " significant reform efforts" in education and accomplishments in medical research. The awards carry substantial funding support.

We found that organizations differ in the types of mechanisms they use to obtain information to make decisions about promising or exemplary programs. While experimental and quasi-experimental evaluation data may be used as one form of information to determine the promise of a given initiative, this type of data is commonly one source of data among many. We have analyzed the different approaches employed by the organizations in our survey, and developed a continuum based on this analysis. We found that many foundations, associations and non profits employ different strategies depending upon the purpose of the information. We use numbers on this continuum only for ease of reading;. They are not intended to connote a judgment about better or worse approaches. The more "methodologically rigorous" may be used in combination with those elsewhere on the continuum and are less frequently employed.

The activities are:

1) relying on key staff for information
2) using external reviewers
3) compiling information that was sent in response to proposals or award applications
4) relying on program developers who are acting for funders
5) employing participatory and/or qualitative methods
6) using data generated from experimental or quasi-experimental demonstration projects

Information gathered from these types of activities is used as evidence for making funding decisions and for highlighting initiatives in the hopes that others will replicate or adapt them.

Each is described briefly below.

(1) **Staff selection** may mark the opening of a new funding priority in which case, the program officer may emerge from the field to be funded - not infrequently having been a former grantee - or be a current staff member who undertakes further training. Customarily the key staff model functions with the assistance of expert opinion in the form of consultants. The latter furnish a link to current knowledge in a field and provide the imprimatur of recognized leaders in an area of interest. The question "are customers involved in the designation of excellence?" would not be seen as applicable by foundations functioning under the key staff model.

(2) **External reviewers** , often in panel form, offer expert vetting of staff opinion and/or guidance in a specialized subject area. One step more formal on the review continuum , these panels of experts assist in the designation of best past practices as well as promising initiatives. An example is provided by a foundation with a long standing commitment to the eradication of a group of diseases; the panel of scientists is able to link their current recommendations to the established record of the foundation as well as to assess the needs of the field. Panel members serve for a specified period of time and their identities are known to applicants for foundation grants, as are presumably the positions of such members on controversial aspects of the direction of future research. (Foundation executives acknowledge that they may attempt to balance the "left and right" attribution of their associates to deflect criticism in this regard .) An organization sponsoring an award may signal its priorities in part through the announced members of its selection panel.

(3) **Self selection as in responses to a general rfp** allows providers ("customers" of a foundation) to offer a rendition of their track record which fits explicit criteria of excellence. The Request for Proposals is given some recognition as being equitable in that all can apply (presuming they know of the opportunity and have the resources to respond in the time allowed). Further, the assertion of excellence and the presence of best practices is summarized by "those who do the work" within the

BEST COPY AVAILABLE

framework of a defined commitment to fund. RFPs can be quite general or specific and may influence the conceptual direction and commitments of non profits . Occasionally a funder will outline a very concrete set of assumptions about excellence and target a small or exclusive group of participants. While a compliment to the providers, this form may circumvent an agency's best judgment. One experienced provider commented after an initiative of this type derailed , "We never would have dreamed this up. ourselves"

(4) **A program developer acts for a funder**. This type of information gathering and selection is specific to foundations. A major established area of the foundation's mission is given prominence through the establishment of a free standing dedicated agency. Its purpose it is to forward that field through research review and strategic funding. The components believed to be critical are derived from literature review and staff expertise. Successful applicants for funding are those that can respond to and utilize the key components which are said to have been "mapped on to". Grants add to accumulating knowledge of the efficacy of the selected components over time. The developer agency is an established staffed organization. While it commends considerable independence in its operations, the program developer does not determine the level of funding. In this sense, the final arbiter is the parent foundation. The publications and professional exchanges fostered by the agency permit and legitimate extended debate about the suggested best practices.

(5) **Employing qualitative - increasingly participatory - evaluations** to obtain information about promising or effective programs or initiatives and to contribute to their improvement. This category includes case study evaluations, participatory and empowerment evaluations that may be commissioned by foundations, associations or non profits . They may be supported by external sources or be conducted by the organizations themselves. The purposes of this type of activity vary. The information may be used to determine promising programs and may be a criteria for program continuation. More often the goal is to highlight and strengthen promising or exemplary programs. Further, these types of evaluations may be promoted to generate information that organizations can use to learn about themselves. That is , there is not simply an "external" audience for the data generated from this type of activity. A strength of this approach is that it can generate empirical data that can be used for decision-making at a variety of levels.

(6) **Using experimental or quasi-experimental** approaches in demonstration projects. Foundations, associations, and non-profits recognize the strength of experimental and quasi-experimental approaches but also note the expense and time constraints associated with these approaches. Foundation people will also share war stories about the unintended negative fall out from some experimental designs. Some foundations, associations and non profits commit funds for this type of activity to generate empirical data to show program effects. One weakness of this approach is that it is not only expensive but when program models are changed even slightly, the cluster evaluation may not capture the reasons behind differences in outcomes.. However, it has the strength of being seen as generating data that can be used in politically charged environments to "prove" the effectiveness of programs.

Section Five: Innovative Evaluation Approaches

We found that the foundations, associations and nonprofits included in our review frequently used innovative approaches to determine promising or exemplary programs. The approach as well as the resulting data are often determined by the mission the organization and the intended target audience. Many organizations promote the use of each of the activities along this continuum.

They believe they recognize the strengths and weaknesses of the different approaches. Determining which is best depends upon the situation. For example, methodologically rigorous evaluations may be very expensive and time-consuming and carry a certain amount of distrust for parts of the non profit community. However, there is acceptance of the importance of such designs to test the effectiveness of demonstration projects.

The *expert panels being developed by OERI* confront several requirements or aspirations familiar to the foundation/nonprofit world. One of them is the desirability of comparative evaluations. Susan Klein writes, '"It is … anticipated that federal agencies and foundations would support the continued evaluation of programs designated as promising and that where feasible, these evaluations would be done in a comparative, but non competitive, fashion using the same performance indicators for programs with similar purposes" (Klein, p 18). Such evaluations could yield cumulative knowledge of good programs. This is both an exciting prospect and one likely to, initially, elicit suspicion. As one experienced funder commented "If the feds are just coming up with a check list, it's a non starter". The situation faced by those seeking to implement common criteria is comparable to that of a membership organization seeking to engender participation in joint standards. One place to look for experience in this regard is Untied Way of America. This long established national organization is aiding its member chapters (those which volunteer) to gradually create a common set of outcome criteria. It is not expected that local chapters will actually compare resulting data until a lengthy experience of the process has taken place. See below United Way of America..

Another need of the expert panels is to gain the commitment of funding sources for their activities. A somewhat parallel experience might be that of those entities which are linked to foundations but relatively free standing. The Carnegie Center for Adolescent Development combines research expertise with selected funding authority. The Harvard Family Research Project, another example, has established links to a group of funders and has demonstrated both research and dissemination capabilities. An OERI panel with the authority to fund follow up evaluations of a number of promising projects might be more viable (and attractive to prospective panel members) than one which would be compelled to seek support for each worthy project after the fact. . To this degree, the panel would become a designator of projects for already committed funders. Such a panel could reasonably be expected to be imaginative in the *type* of evaluation it advocated as well as the high caliber of the project to be assessed An example of such alertness is furnished by Innovation Network, a non profit evaluator in Washington DC , which is currently field testing a Tool Box of evaluation aid to be implemented via the internet. The pre test will determine how much cost reduction might be achieved due to reduced face to face contact between evaluator and program and what quality of service can be achieved.

Acting as members of one of the Expert Panels may be problematic for foundation professionals, if asked , because of their disinclination to appear to be endorsing specific grants prior to a vote of their Boards of Trustees. However such service might be more likely to occur if the affinity group, Grantmakers Evaluation Network, of the Council on Foundations was asked to nominate two or three of its members for a period of service. Another possibility would be to ask for nominations for the panels from the Non-Profit and Foundation TIG of the American Evaluation Association.

Another possible parallel between the work of the Panels and the efforts surveyed for this paper is the significance of the strength and enthusiasm provided by committed participants. . For Parents as Teachers and Impact II, Teachers Network, for example, finding best practices was an act of faith in members' roles and importance. Hence, the organizers could count on contributions being made. If comparative, noncompetitive evaluations are to be instituted by the OERI Panels, such camaraderie among participants will be needed. It may be noted that both of these two programs utilize group norms that guard against strains. E.g., Impact II, Teachers Network describe their project selections as being "noncontroversial" and Parents as Teachers requires a shared training period..

The issue of "non competitive comparisons" raises familiar issues for national foundations. The Kellogg Foundation, for example, in working with cluster evaluations and the Lily Endowment with an innovative series of grants to improve school counseling have devoted considerable amounts of time and resources to this end. They have employed technical support and assistance and nontraditional evaluation procedures to involve project participants. They have also gained the reputation of trusting grantees to both give and take in the on-going exchange.

The OERI article cited earlier makes a further important point, "…another challenge for the System of Expert Panels is to develop new ways to connect dissemination and evaluation incentives . " The Robert Woods Johnson Foundation has been particularly effective in integrating the work of communication professionals with program creators and administrators. Dissemination of information about upcoming funding opportunities is very widely disseminated in ways that aid applicants to make their best case. Further, explanations about a given grant opportunity are linked to general questions of moment for the field; readers can feel they have gained from reading an rfp even if they don't undertake to respond. Past and present grantees are lauded but their work is also analyzed in detail. Thatis, a project can expect to see its project outlined with both generosity and truth as in the case of patient choice in medical treatment. (See below). The Harvard Family Research Project, following the lead of the Harvard Education Newsletter, has influenced professional debate by mounting a truly interesting newsletter.

As a summary to this paper, we would note that all of the organizations in our study rely on staff and expert panels for many of the decisions they make regarding promising programs. In addition, those organizations that appeared most confident of their conclusions used multiple methods for obtaining information about them. Participatory and empowerment evaluations not only helped identify promising models but also generated information that was judged useful to the programs.

Appendix I

SUMMARY OF THE WAYS IN WHICH SELECTED FOUNDATIONS, ASSOCIATIONS AND NON PROFITS SET STANDARDS AND DETERMINE PROMISING PRACTICES

DEWITT-WALLACE READERS DIGEST FUND

The DeWitt-Wallace Readers Digest Fund uses a variety of strategies to determine best practices. The Fund's grant-making strategy is guided by its mission to improve quality education and career development for all persons. Currently, the Fund is focusing on three areas: schools, community-based organizations, and school/community partnerships. The standards that are used in developing and applying the Fund's policy to meet this mission involve a number of steps. The Fund surveys the field to determine the best available knowledge about specific activities that could make and a difference and that are matched with the Fund's mission. The Fund then examines the research and commissions papers, especially from those who are most knowledgeable in the field. This research is translated into strategies that will meet the Fund's goals. Experts in the field, including practitioners, are asked about programs they believe are outstanding , proactive and viable. Funding decisions are made based on nominations and with consideration of the Fund's mission. The Fund has standards regarding the evaluation and assessment of the activities that are awarded grants . Based on nominations and with consideration of the Fund's mission. The Fund has standards regarding the evaluation and assessment of the activities for all persons. Grantees vary in the levels of evaluation and assessment they engage in, depending on the nature of the grant. All grantees provide documentation of their program so that it can be replicated in the future. Grantees that have the capacity for self-assessment are asked to provide such documentation. This information can be both useful to the grantees in improving programs and to the Fund. Some grantees hire external evaluators to conduct the evaluation. Finally, the most rigorous evaluations are those commissioned by the Fund. These tend to be for very large grants that are seen mostly as research and demonstration efforts.

HARVARD FAMILY RESEARCH PROJECT AND RESULTS-BASED ACCOUNTABILITY

Harvard Family Research Project (HFRP) is a programs that support families and communities to help children reach their fullest potential. It seeks "to capture the voices of those who work intimately with families across a broad range of settings - schools, social service agencies, youth-serving organizations, daycare centers, and various types of nonprofit entities". It is, in this sense, very practitioner based. HFRP researches preventive, comprehensive, collaborative, and family-focused early childhood services and professional training, and helps translate those findings into action.

Of the many on-going HFRP activities that apply standards in evaluating and disseminating findings, one example is the Results-Based Accountability Project (RBA), funded by the Pew Charitable Trusts and the Ford Foundation. Great public emphasis has been placed on such accountability as a part of recent major changes in governmental support of families. Practitioners, in the case of this project, are those public servants and policy makers who must implement major new systems and satisfy their constituencies concerning outcomes of the new programs. RBA is intended to support and build upon efforts going forward in state and local communities to develop accountability systems for child and family services. The Project is examining results-based accountability efforts --which include efforts to systematically articulate strategic plans, specify goals and targets, and report indicator data on progress toward achieving the targets. States, facing substantial challenges in implementing new approaches, have expressed interest in learning from each other. The Project is examining eight states in depth to delineate lessons and promising approaches in developing and implementing these innovative systems. The eight states (MN, OR, VT, NC, OH, Iowa, GA and FL) were selected in a multi step process. First, significant constituent groups such as the Council of Chief State School Officers and the National Governors' Association, were invited to nominate states. All states were interviewed and a team of four HFRP researchers narrowed the group to 18. Documents the states provided were then reviewed and additional follow up interviews were conducted. At this point, practitioner knowledge was folded back into the design of the RBA project by means of asking representatives of the 18 states to comment on what information should be sought and what information would have been useful to them as they were designing their efforts. As a result of this aspect of the project design, the criteria of selection of the final eight states included political as well as demographic factors.

The eight states taking part in the in depth case studies represent not only a range of models but also are in different developmental phases of design and implementation of their efforts. Some states such as MN and OR have already experienced evaluations of their efforts from a variety of sources. (In MN the state legislative auditor evaluated and in OR multiple reviews have been conducted, one by the Annie E Casey Foundation, another by the Kennedy School's Innovations in Government Project, others by the National Performance Review.) Other states among the eight have conducted formal evaluations of their efforts specifically as they related to child and family services (examples would be the Frank Porter Graham evaluation and the Coopers and Lynbrand evaluation of the Smart Start Initiative in NC.) All of the evaluation efforts being carried out are summarized in the forthcoming case studies. The evaluations range widely in approach, representing all the types described earlier in this paper.

As in the selection process, the Project applies standards throughout its data collection and dissemination activities. HFRP obtains information through structured interviews, document reviews, and site visits. The goal is to ascertain the challenges and promising approaches people are encountering, and to continually ascertain the information needs of those involved in these efforts. The standards applied in the information dissemination strategy of the Project involve the requirement that each such activity meet one or more of the following goals: 1) to share, in a timely manner, HFRP's analysis of issues regarding results-based accountability; 2) to highlight the efforts of states who are planning and undertaking the development of these initiatives; 3) to facilitate communication among those currently involved in these efforts; and 4) to act as a clearinghouse of resources and information for those interested in results-based accountability. Examples of information dissemination activities include the HFRP quarterly newsletter entitled The Evaluation Exchange which is designed to both provide HFRP analysis of the issues in a timely manner and to highlight the work of those involved in planning and developing these activities. The newsletter has carried discussions with the state actors and anaylsis of the unfolding projects and its results. Another example is the technical assistance HFRP is providing to states and localities by referring callers to those who have encountered similar challenges and by providing specific advice based on HFRP analysis of the issues. The technical assistance also facilitates communication among states and allows HFRP to act as a clearinghouse of information.

REPLICATION AND PROGRAM STRATEGIES, INC.

Replication and Program Strategies engages in a number of activities to promote replication of promising and effective programs. This organization provides technical assistance, analysis , and education, to increase use of replication of effective programs strategy. It also assists programs in expanding and deepening knowledge about replication and related strategies for extending impact of worthwhile programs. Further, it conducts feasibility assessments to determine if a program is ready for replication. Replication and Program Strategies provides technical assistance to organizations through a comprehensive strategic planning process. The goal is to provide mid-course development assistance. The primary strategies the organization uses is promoting the strategic planning process, problem solving and designing skills to help ongoing replication pinpoint problems, needs and opportunities. It then works to devise strategies for improvement

ROBERT WOOD JOHNSON FOUNDATION AND PATIENT CARE

The Robert Wood Johnson Foundation determines best practice through setting standards and assessing programs against these standards. One of the nation's largest foundations with over $4 billion in assets and a specialized funding area, RWJ begins its standard setting with the selection of Foundation staff. Their professional familiarity with major health issues permits extensive internal review of a possible new activity or subfield . An example of the development of an issue area is the Foundation experience with the choices involved in the selection of medical treatment for the hospitalized seriously ill. Consultants and expert panels were used by the Foundation after internal review by staff to specify the accomplishments and outstanding research questions of the specified area. A sophisticated communication staff is involved early in formulating outreach strategies that inform researchers and practitioners of the availability of funding in the subfield. .Learnings from initial grants and their contributions to the Foundation's evolving standards are also disseminated. In the case of the determination of patient care for the seriously ill, one .standard was the involvement of patients in treatment selection including the choice of having no treatment. It became clear in the course of a series of major grants, that a major expectation of the Foundation's was not confirmed: When physicians gained knowledge of consumers' preferences in final care, they did not commonly respond by implementing those preferences. The Foundation, given its familiarity with the field and its continuity of interest, moved, after review, to the pursuit of other means of adding patient involvement in final care decisions. RWJ sought wide discussion among professionals and various publics. Rather than retreating or downplaying the failure of their original assumption, the Foundation invested energy and resources in seeking to find the causes and alternative approaches. Publications and conferences allowed for the consideration of a range of responses, including those of professional associations of doctors. In this way advocacy was directed at the overarching structure of professional understanding and belief .

UNITED WAY OF AMERICA

The United Way is an organization evolving its position and capacity on assessment and replication. The potential for impact through such changes in its procedures is considerable . A vehicle for the charitable expression of local business and institutional concerns, United Way county and city affiliates play a pivotal role in many communities, supporting social service organizations and influencing the decisions of other funders. Currently, member donors face demanding choices for their charitable contributions and there is increased demand for information about best practices and the effectiveness of programs. For the national office, meeting such demands means enhancing the ability of United Way chapters to "demonstrate the value that United Ways add to donor funds by aggregating them …and ensuring that funded programs are effective." [234] Thus, there is a clear organizational imperative, as well as intellectual commitment, in the search for usable outcome measures. Early steps included a review of the literature and the design of member symposiums where methods were discussed. Local chapters highlighted their own procedures but there was no intent or pressure to seek a consensus.. The meeting and exchange stage was followed by an analysis by the national research staff of the probable steps, or increments of development, a local chapter

---

[2] "Measuring Program Outcomes: A Practical Approach", United Way of America , 1996.

[3]

[4]

could expect to move through as its practices in this area evolve. The procedure followed by the national office has thus been to draw the attention of United Way affiliates to what others have done in order to establish clear standards and to alert them to what they can expect to occur if they choose to pursue outcome measurement..

The United Way of America focus on effectiveness is described in a 1996 publication "Measuring Program Outcomes; A Practical Approach" which describes goal articulation, the creation of indicators and the reporting of data. It is stressed throughout the publication ,that, although examples may be offered, the outcomes to be used as targets are those specified by the local UW . The local UWs have not been asked to sign on to the implementation of a common national evaluation model. The power of decision rests with the community volunteers who staff distribution committees. However some comparison among the chapters is anticipated. In this, as in other efforts considering best practices, those attempting change thrive on opportunities to hear from peers. Data from a large pool of service providers should become available over time to a national research staff committed to documenting and replicating best practices. Currently, six local chapters , those judged to be the furthest along in the pursuit of outcome measurement, are receiving support to evolve a set of criteria in stated subject areas , such as services for senior citizens and youth at risk.

IMPACT II, THE TEACHERS NETWORK

Established in 1979 in New York City, the non profit, Impact II The Teachers Network, has grown to be a national organization which disseminates descriptions of teaching programs and furnishes grants to teachers. As its Executive Director Ellen Meyers commented, " We were pushed by our own constituents" to expand. The values being lauded by the Network include cost effectiveness and staff development with a goal of improved curriculum. Impact II involves practitioners directly in the designation of best practices . Teachers submit curricula they have developed and are recognized by small grants. The involvement of teachers at all levels of the process defines the effort and is a source of cohesion and camaraderie .

There are several aspects of its operations which make Impact II instructive:

- The selection process is carried out by peers. Practitioners are central to the designation of best practices, as well as being the submitters. Classroom teachers self select to submit a unit of their practice; other teachers then judge which proposals will receive small grants and be actively disseminated for adoption by others. While the selection committees do contain some administrators and "friends" of the effort and union representatives, the role of enthusiastic classroom teachers is the key one.
- The descriptions of the awarded units is written by the teachers and the worth of given ideas is calculated over time via the number of adoptions made by other teachers . In this way practitioners are the fulcrum of the creation and lauding of what works in teaching.
- Advocacy for classroom teachers inheres in all components of the Network's activities. Symbolic of this orientation were the public objections the organization made to a presidential education summit in 1989 which "did not invite a classroom teacher".
- The grant selection process is not viewed by Network personnel as involving controversy. Best practices in this sense are simply the reality of what good teachers do.
- The commitment of the Network to recognize and reward teachers and to disseminate information about their classroom programs also includes a commitment to encourage continuing profession development. Regional committees of experienced Network teachers are available to support those beginning the application process. Classroom visits between Network members are encouraged and sometimes paid for. There are various forms of assistance to those who choose to adapt a curriculum unit including " adapter grants".
- The evaluation of the program does not report on student gains or perceptions but rather views the opinions and the enthusiasm of teachers for their work as the key variables by which to assess the success of the Network..

Some independent schools have been involved in the Network and , in a limited number of cases , a school system, as opposed to an individual, has joined the Network. But the focus remains with individual 'self starter' teachers submitting peer reviewed curricula units and explaining to professional colleagues why the unit was successful.. The Network views part of its task as creating a social network among innovative teachers and identifying "the deepest concerns of teachers" as well as creating better individual teaching units.

Evaluation of this program has been of long duration. Dale Mann, Teachers College faculty member and well known commentator, carried out data collection in 1981 and again in 1992-3. He found that after 10 years "the main effect of Impact II's networking remain(s) strongly in place". Among the effects he reported were greater confidence experienced by participants in their professional role and gratitude for contact with peers.. Mann notes that teachers commonly see themselves as isolated and manipulated. In his opinion, national efforts and conferences characteristically talk down to teachers, increasing an "infantalization" within the profession..

Funding for the Network has come from a variety of foundations and (formerly) from the federal government's Diffusion Network.

THE CARNEGIE CENTER FOR ADOLESCENT DEVELOPMENT

The Center carries the name and long standing commitment of a single prestigious funder. Its cumulative grantmaking decisions and findings are influential and closely linked to the thinking of its parent organization. It is however freestanding; its employees are staff members of the Center not of the Carnegie Corp. Grant proposals received by the Center are reviewed and judged by Center staff and consultants. It commits its efforts to increasing professional and public understanding of adolescent needs and potential. It seeks best practices in the field by establishing a set of criteria and using these to map on highly relevant proposals. Consumer impact is experienced through the high degree of interaction with leading programs and providers.

THE YOUTH DEVELOPMENT INSTITUTE OF THE FUND FOR THE CITY OF NEW YORK

an operating foundation encountered late in the research for this paper, provides a different model. Its search for best practices is carried out through a selected group of service providers. Their contributions include on-going roles in the research and support activities of the Institute. Their shared self examination has yielded conclusions about best practices and they have welcomed assessment teams into their own programs to clarify and specify these practices. The teams are lead by agency practitioners and draw membership from youth participants and Institute staff.

The vehicle for these activities is the Institute's Network for Youth Development, which is a peer network of youth organizations "promoting youth development as a field of practice and mastery and committed to accountability and authentic assessment"[5]. A primary purpose is to strengthen practitioners. Comprised of nine well known youth agencies selected through a competitive request for partners, the Network, which benefits from the active involvement of Institute staff, has committed itself both to developing a consensus on positive outcomes to be achieved by young people and to delineating core competencies needed by youth workers. The Network offers its conclusions for the field of youth development to consider.

First, the Network identified *areas of competency* for youth participating in its members' programs. Designated under headings of Creative, Cognitive, Civic, Employability, Personal, Physical Health, Mental Health, and Social Competencies, the stated areas are both general and demanding. Demonstrable outcomes, which may be subjective or objective, are addressed in each of the areas of competency. For example, under Employability Competency, skills listed include "Dressing and grooming for an interview" and "Being able to manipulate without challenging". Mental Health Competency includes "Development of positive self-image" and "Maintaining physically healthy life style".

---

[5] "The Handbook of Positive Youth Outcomes", The Youth Development Institute, The Fund for the City of New York, undated.

These attainments are described as "progressive competencies". Thatis, youth will develop them at different rates and demonstrate them in different ways. The competencies are the overall results of programs operating under, what the Network believes to be, a successful youth development framework. Thatis, the Network takes responsibility for specifying and holding up to view, clear outcomes for young participants in its members' programs. In developing its approach, the Network members asked themselves, and colleagues elsewhere, questions which have sometimes been avoided in other efforts. They included:

- "Should youth development agencies hold themselves responsible for progress on the attainment of these outcomes by the young people who are in their programs?"
  - "Are these outcomes that all youth development programs, regardless of service content or funding source, should aim at for their participants?"
- "Are we willing to be held accountable for progress towards these outcomes in our young people?"

Assessment was from the beginning of the Network's activities an important pledge to the field. An Outcomes Committee was created to identify indicators of the positive outcomes. This brought about refinement of the consensus on the competency areas . One result was that the staff was asked to regroup the competencies in to crosscutting skills areas such as the ability to negotiate. In this and other Institute activities, past research on youth development has been funneled through members' practices and experience in several ways.

With the assistance of a professional evaluator and the Institute staff, the agencies developed an assessment protocol. Its stated purpose was to field test the work on best practices and to reveal to what extent, the Network's members' programs match them. Assessment teams were made up of 2 peers, thatis staff from other agencies, 1 lead person from the agency being assessed, two young people and two Institute staff. Review practices included document review, site visits, interviews and observations. The team identified the strengths and needs for improvement of the programs.

The activities of the Institute benefit from the established links of the Fund with local government and public policy makers. Wide dissemination is being undertaken of the implications of Institute findings for the field. In concert with the tone and commitment of the Network, their publications are designed and issued by a youth enterprise.

INTERNATIONAL YOUTH FOUNDATION

A unique funding agency, the International Youth Foundation (IYF) is linked to six national or regional foundations from a variety of countries. The ten member Board of Directors is drawn from eight nations. While sharing a commitment to youth work, the member foundations remain independent. The shared concern of the members is the positive development of children and youth ages 5 to 20. The IYF reports that 70% of its income comes from foundations. W.K. Kellogg Foundation in the USA has been an active supporter as has the C.S. Mott Foundation. The assets of IYF were reported to be $30.6 million in 1995.

The mandate of the Foundation is to enable successful programs to share the ways in which they are helping young people world wide. Through funding, technical support, peer interchange and a skilled dissemination effort, the knowledge of excellent programs is nurtured. Another goal of the Foundation's work is to contribute to an understanding of - and hence support for - the youth development field on the part of a broad public. A tone of great urgency pervades the Foundation's published materials. It is essential, they suggest, that support and empathy be achieved for young people facing extraordinary challenges..

Submitted programs are reviewed by experts utilizing seventeen criteria for youth program effectiveness. Some of the criteria are demographic - e.g., "serves youth between 5 and twenty years old". Others are organizational, e.g., "shows financial and managerial capacity and displays cost effective means of implementation". The Foundation stresses a strong commitment to age appropriate activities and to promoting character . The projects supported by the IYF must in addition, show effectiveness, sufficient scale and sustainability. As the Foundation's Director of US Programs, Karen Pittman, notes, "We must break out of the 'demonstration mode'".

Evaluation of its efforts is a prominent concern. Pittman writing in 1995 said, "We are woefully behind…in conceptualizing what it is about the programs that works. We have effective programs. What we do not have is a solid system for assessing, improving, sustaining, and expanding their impact." Since that time, the IYF has brought together colleagues and evaluators from around the world to grapple with how best to assess what they do and how to involve participants compellingly in that aspect of their work.

****

The conclusions arrived at in the forgoing paper came from a review of fourteen entities—foundations, membership organizations and non profit agencies. In most cases, a single, specific program was considered rather than all activities of the organization.

The intent was to find revealing methods of arriving at the designation "excellent program" in order to aid OERI in fulfilling a legislative requirement. There was a particular concern to document the involvement of the public in the process of such designation and to ascertain the presence of quality control measures employed.. However the realities of the operations and histories of the diverse entities examined meant that very few have formal customer involvement of the type advocated for public agencies. The frequent engagement of former and current clients and grantees is seen by participants as influencing the selections made by them. Standards applied to practices in order to cite them as exemplary did not have the set dimensions of specified time and character implied in the guidelines for the commissioned papers. The conclusions of the fourteen organizations were most often the result of expert review and staff exchange. Foundations may seek to review their fulfillment of goals for subject areas or for the foundation's whole portfolio and thereby come to conclusions in retrospect about best practices

Foundation may also "scan the field" by establishing semi independent entities which not only make funding decisions but determine criteria of selection for an entire area. An example is the Carnegie Council on Adolescent Development. A third way in which foundations specify standards is through the underwriting of award programs such as the Dana Awards. The discussion and research accompanying these selections suggest criteria for the Foundation and for others.

In the case of youth development agencies , the criteria are influenced by the stance of many workers in the field who seek a non segmented approach to services to youth in their communities. For the membership organizations, the search is for a method which local units can fully own and adapt to their own settings. . For the research and program implementation groups, the imperative is to influence action by adding to knowledge while at the same time responding to the needs of funding partners.

The presence of quality control for these organizations is represented by the use of evaluations in many of the agencies listed . An agency of long standing such as Teacher Network IMPACT II may have the benefit of repeated evaluations and can demonstrate gains and the retention of achievements.. However the evaluations utilized by the programs or projects are by no means similar in approach or demandingness.

What can be gained from this review is an intimation of the diversity of approaches utilized and an acknowledgment of a common desire to learn from others. In the case of the Youth International Foundation, for example, all actions draw on the need to share and learn and to build in those supports which will allow transfers to take place.

APPENDIX II

. The following list contains summary information on the individual entrants and also comments briefly on their uniqueness. Contact information is provided.

Entity: Robert Wood Johnson Foundation
Type: Foundation
Subject area: Funding of health care and policy initiatives to improve health of all Americans
Included in this review: Health care initiative for hospitalized terminally ill
Process: Staff review, expert panel
Comment: The Foundation uses its continuity and knowledge of its field to revisit assumptions and evaluate their ramifications. In one major follow up effort an assumption governing a major grant with teaching hospitals was found to have been incorrect. That was that furnishing physicians with information about patients' wishes would influence their practice. The follow up activities supported by the Foundation were extensive and subsequently shaped professional awareness. The pursuit of excellence in this way directly impacted both funding selection by the Foundation and awareness within the health care profession.
Contact: Robert Graham Hughes, RWJ, Route 1 and College Road East, PO. Box 2316, Princeton, NJ 08543, (609) 452 8701

Entity: IMPACT II The Teachers Network
Type: Nonprofit coordinating group
Subject area: Support of classroom teachers

Included in this review: Small grants program
Process: Submission of applications from classroom teachers who may have had assistance from regional coordinating groups.
Review by network of peer commentators and central staff, and advisors
Comment: National network with successful history of eliciting practices held to be outstanding by practitioners. Dissemination among classroom teachers is a major aspect of mission.. Organization focuses on intellectual and psychological experience of teachers. It does not quantify successful outcomes for students.

Contact: Ellen Meyers, Director, Impact II, 285 West Broadway, NY, NY 10013 (212) 966 5582

Entity: United Way of America

Type: Membership association
Subject area: Funding of social service programs by local member chapters

Included in this review: Accountability Impact Task Force and implementation
Process: Review of best practices by local UWs, selection of exemplary instances by staff and national panel of experts.
Comment: National office is mindful of need to gain members' participation

in building accountability practices while affirming local control and initiative. In service training is a goal.

Contact: Martha Taylor Greenway, Senior Director, Effective Practices and Measuring Impact, United Way of America (703) 683 7830, martha.greenway@uwa.unitedway.org
To order agency manual and training materials: Sales Service America (800) 772-0008

Entity: Harvard Family Research Project

Type: Nonprofit research and program creation agency
Subject area: Family and children
Included in this review: Results Based Accountability Project
Process: Staff and experts from field set criteria for both selection of states to be cited and researched and for the dissemination practices to be followed

Comment: Influence of project results from high standing of HFRP staff and the ongoing support offered funders and others in examining best practices.

Contact: Heather B. Weiss, Director, HFRP, Harvard Graduate School of Education, Longfellow Hall, Appian Way, Cambridge, MA 02138 (617) 495 9108 Fax (617) 495-8594.

Entity: Replication and Program Strategies, Inc.

Type: Research and support non profit organization
Subject area: Expansion of replication for excellent social service and educational programs
Included in this review: Overall program
Process: Review of field by experts
Comment: Bears both the leverage and the liability of being direct outcome of foundation need and interest. Its creation was linked to a major community wide effort to examine replication. It seeks to expand knowledge about strategies for extending the impact of good programs. The agency provides customized technical assistance in all aspects of replication to funders and program operators. It has not been clear to all observers however if research activities or direct service is the chief goal.
Contact: David Racine, President Replication & Program Strategies, Inc., One Commerce Square, 2005 Market Street, Suite 900, Philadelphia, PA 19103, (215) 557 4483 FAX 557 4485

Entity: MDC, Inc.

Type: Nonprofit research organizations and service provider
Subject area: Support of educational improvement
Included in this review: Indiana School Guidance and Counseling Leadership Project
Process: Cross visitation among participants is major commitment of the project and long term learning through the use of extended observation and reflection practices.
Comment: Resists categorizing participants' work as "best practices" since affirmation of community participants is major goal. Long term evaluation and technical assistance contributes to peer learning and shaping of next steps.

Contact: David Dotson, MDC, Inc. P.O. Box 17268, Chapel Hill NC

Entity: International Youth Foundation
Type: Foundation linking foundations globally
Subject area: Youth development
Included in this review: YouthNet International which is made up of programs from over 30 countries.

Goal is to strengthen and expand through mutual support

Process: while involving youth in all aspects of the work. Funding, technical asistance and joint conferences are utilized.
Comment: YouthNet is currently expanding its activities in the Untied States and Africa, most of its granting has been divided among Latin America , Europe and Asia. The organization has evolved clear strategies for promoting excellence including creating the global network and raising awareness of children and youth issues. The organization also stresses a commitment to the local nature of its member foundations. Recognizing the isolation experienced by many good programs, the Network has promoted evaluation and dissemination skills among its members. Karen Pittman, Director of US Programs wrote "A key piece of what makes programs work is having the right mix between service delivery, youth empowerment, community involvement and issue advocacy" reflecting the organizational commitment to a holistic perspective.

Contact: Marita Irby, 67 West Michigan Ave., Suite 608, Battle Creek, MI 49017 (301) 270 4801 Fax 87 7735

Entity: Carnegie Council on Adolescent Development

Type: Research and implementation organization attached to single funder, influencing policy of parent organization

Subject area: Early Adolescence
Included in this review: General operations
Process: Components of excellence are delineated by expert consultation.

Select criteria drawn from research are applied in judging applications for funding.

Comment: The Council has gained recognition through setting standards in its area. By so doing it acts both as an independent funder and an influence on its parent organization and the field generally

Contact: Former Director interviewed, Carnegie Council on Adolescent Development 2400 N Street, NW, 6th Floor, Washington DC 20037, (202) 429 7979

Entity: The Charles A. Dana Foundation, Inc.

Type: Foundation

Subject area: Education and medical research

Included in this review: Charles A. Dana Award for Pioneering Achievements in Education

Process: Submissions for the awards are sought in a variety of ways. Selection is made by group of experts with some carry over from past years. Finalists are provided with both financial and dissemination assistance. A much discussed aspiration of the Award is the manner and success of replication and adaptation of the award winners' programs.

Comment: Leaders in the field of educational change such as Parents as Teachers have been prominent among the Dana winners. The achievement of the award brings them to the attention of other funders and the media. Extensive written commentaries give in depth understanding of their achievements

Contact: The Charles A. Dana Foundation, 745 Fifth Ave., Suite 700, NY 1051 (212) 223 4040

Entity: DeWitt Wallace Reader's Digest Foundation

Type: Foundation

Subject area: Education, youth

Included in this review: Grantmaking selection policy

Process: Expert and staff review building on established evaluation procedures.

Comment: The Foundation has experienced major restructuring in recent years with an important commitment to the area of evaluation and assessment. The prominence of its work in education has influenced other funders.

Contact: Andrew Fisher
2 Park Ave.,. 23rd Floor, NY 10016
(212) 251 9700

Entity: W.K. Kellogg Foundation

Type: Foundation

Subject area: Youth, education, community development and leadership development
Included in this review: Middle South initiative
Process: Foundation staff devoted development time to meeting with present and former grantees and gaining expert commentary on needs of an area which was believed to be underserved. See grants were given to develop and draw upon local community building efforts.
Comment: The Foundation has used both its evaluation staff and contacts with smaller foundations within the geographic area to arrive at needs and goals. Cluster evaluation will be utilized through the life of the initiative.
Contact: Ricardo Millett, One Michigan Ave. East, Battle Creek, MI 49017 (616) 968 1611 FAX (616) 968 0413

Entity: The JM Foundation

Type: Foundation
Subject area: Rehabilitation of people with disabilities and entrepreneurship
Included in this review: Search for Excellence in Vocational Programs for people with disabilities and the initiation of National Results Council

Process: Foundation's Search for Excellence (award for which rehabilitation professionals submit applications) revealed sizable gaps in standards and utilization of means to efficiency. Thatis the award's juries found some agencies with very efficient delivery of services while other lagged far behind on the slected indicators. The Foundation, drawing on this experience, sponsored the National Results Council. The Council articulated standards derived from the Award's experience and seeks to set up a credentializing process specifying steps by which results can be achieved. The goal is efficiency as well as excellence of service delivery
Comment: The engagement of other foundations is sought so that the credential process can gain recognition. Steps include organizational and curriculum changes.

Contact: Chris Olander, The JM Foundation
60 East 42nd St., Suite 1651
NY 10165 (212) 687-7735

Entity: Public Private Ventures

Type: Nonprofit research and program development organization
Subject area: Disadvantaged youth with special interest in their participation in work force

Included in this review: Big Brothers/Big Sisters of America evaluation study
Process: Long term evaluation of this program lead to delineation of what

works in a mentoring program.

Comment: The criteria arrived at and publicized were of a type which can be adopted by others. Both the substance and duration of the evaluation and its dissemination were significant. Mentoring has been an approach of great emotional appeal but, prior to the Public Private evaluation, relatively little clear documentation.

Contact: Gary Walker, Public/Private Ventures, 399 Market Street, Phila., PA 19106 (215)

Entity: The Youth Development Institute of the Fund for the City of New York

Type: Foundation linked non profit research and advocacy organization
Subject area: Youth development
Included in this review: Networks for Youth Development
Process: Best practices, the presence of which would indicate an excellent program, were arrived at through examination of the work of members of the Network. The youth agencies (9 service providers and the Fund) make up a working, advocating and professional development effort. The standards applied to best practices are disseminated on the basis of the long time experience and respect won by the member agencies.

Comment: Youth development professionals appear to increasingly agree on characteristics of best practice (See International Youth Foundation) and to be gaining the support of funders for their utilization.

Contact: Michele Cahill and Linda Pitts, Fund for the City of New York, 121 6th Ave., 6th floor, NY, NY 10013, (212) 925-6675 FAX 925-5675

J

Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas

John Luczak and Joan Ruskus, SRI International
U.S. Department of Education
Office of Educational Research and Improvement
July 10, 1997

Table of Contents

Background and Organization 1

Identifying Review Efforts 1

Analysis and Overview of Review Efforts 2

Vocational Education Review Efforts

- Center on Education and Work's "National Consortium for Product Quality" 6
- National Youth Employment Coalition's (NYEC) "Promising and Effective Practices Network (PEPNet)" 9
- Southern Regional Education Board's (SREB) "High Schools That Work" Program 13
- Transition Research Institute's "Taxonomy for Transition Programming" 17
- Center for Occupational Research and Development's (CORD) "National Coalition of Advanced Technology Centers (NCATC)" 20

Special Education Review Efforts

- National Center for Research in Vocational Education's (NCRVE) "Exemplary Career Guidance and Counseling Programs" 23
- Office of Special Education Program's (OSEP) "Screening Promising Practices for Adoption" 27
- Idaho's "Special Education Exemplary Programs" 29

Educational Equity Review Efforts

- Office for Civil Rights (OCR) "Promising Practices and Programs" 32
- American Association of University Women's (AAUW) "Growing Smart" Report 34

Early Childhood Education Review Effort

- National Association for the Education of Young Children's (NAEYC) "National Academy of Early Childhood Programs" 37

Postsecondary Education Review Effort

- Fund for the Improvement of Postsecondary Education's (FIPSE) "Lessons Learned" ... 41

APPENDICES

Appendix A: Sample Profiles

- Transition Research Institute's "Bridges...from School to Work" program
- PEPNet's "Arizona CALL-A-TEEN" program
- FIPSE's "City University of New York" program

Appendix B: Sample Applications

- Transition Research Institute's "Effective Transition Practices Nomination Form"
- NCRVE's "1996 Exemplary Career Guidance and Counseling Program Search"
- PEPNet's Application

Appendix C: Sample Rating Forms

- Idaho's "Special Education Exemplary Programs" Scoring Process
- NCRVE's Rating Form
- OSEP's "Screening Promising Practices for Adoption" Checklist

Background and Organization

The 1994 reauthorization of the Office of Educational Research and Improvement (OERI) in the U.S. Department of Education mandated OERI to: 1) develop standards to designate promising and exemplary products, programs, and practices; 2) establish a system of expert panels to make recommendations on the above designations; and 3) coordinate dissemination activities and programs within the Department and other government agencies, as well as with associations, state education agencies, and local school districts.

OERI has been charged with helping educators make more informed decisions about promising and exemplary products, programs, and practices as they attempt to implement these improvement efforts. The office has published a set of standards for public comment (June 1996) and is currently pilot testing two expert panels in the areas of gender equity and mathematics and science education. OERI has begun conducting research on existing review efforts that attempt to identify and share the best of what is available with consumers. This paper was commissioned to review efforts in a wide variety of education organizations and topic areas, excluding review efforts of mathematics and science programs.

Twelve such review efforts are described and analyzed in this paper. Key aspects and cross-program themes are discussed in the analysis and overview section that immediately precedes the individual review descriptions/profiles, which make up the majority of this paper. **These profiles frequently borrow direct passages of text from written materials received from the review efforts.**

The review efforts focused on a wide variety of topics in education: five reviews evaluated vocational education programs or practices, three were special education reviews, two concentrated on educational equity, one targeted early childhood education, and one other review effort focused on postsecondary education. Each profile includes available contact information, funding information, the purpose of the review effort, the type of resource reviewed, the criteria used, how decisions were made, information about the reviewers, and any post-review efforts. The Appendices include a selection of program profiles, sample applications, and sample rating forms from the 12 review efforts.

Identifying Review Efforts

The process to identify review efforts was not systematic in a research sense. OERI made initial suggestions of certain review efforts and provided names of important contacts who were able to identify other promising efforts (similar to snowball sampling). The principal criteria for selection was representing a wide range of review efforts with as little duplication as possible.

Analysis and Overview of Review Efforts

The cross-program analysis of the 12 review efforts is divided into six sections: types of resources reviewed, types of review efforts, evaluation/effectiveness criteria, evidence of success, incentives for programs, and lessons learned for "promising" and "exemplary" standards and expert panels.

Types of Resources Reviewed

Most of the efforts (two-thirds) examined reviewed programs rather than products or practices. The programs reviewed were diverse and included the following: advanced technology centers, transition programs, high schools combining challenging academic courses and modern vocational studies, career guidance and counseling programs, state-level special education programs, girls' achievement and healthy development strategies/programs, early childhood programs, and higher education reform programs.

Of the remaining efforts, one analyzed School-to-Work instructional products while another effort evaluated youth development/employment practices. Another review effort created a generic screening process for use with any type of product, practice, or program. And finally, one review process involved five or six separate review efforts and therefore examined all three types of resources.

Types of Review Efforts

There was significant diversity among the 12 review efforts in terms of approach used to review materials. The approach was often determined or influenced by 1) the review process itself, and 2) the number of programs or practices that needed to be reviewed.

The following is a mini-typology (with some overlap) of the different review effort approaches:

- Formal rating system used with point totals in review process (3 review efforts)
- Program applications reviewed by experts (2), or reviewed by experts after a first round of review by project staff (3)
- Applications reviewed by team of staff with semi-structured process in place (4)
- Programs or practices reviewed informally by small team or single staff member (1)
- Generic screening process developed for a diverse group of users (1)

Almost all of the review efforts had a set number of program applications to evaluate. Many reviewed materials sent to them through a call for applications or nominations, while others received a consistent number of membership applications each year. Only two efforts conducted their own searches for promising practices and programs. The search for girls' achievement and healthy development strategies/programs involved a review of more than 500 studies and reports. The other effort involved five or six separate reviews conducted by staff who each put together (in a particular subject area) a list of at least ten promising programs from the multitude that existed.

Evaluation/Effectiveness Criteria

Almost all of the review efforts had evaluation or effectiveness criteria and used them to evaluate prospective programs or practices. The evaluation/effectiveness criteria for programs obviously differed depending on the subject area and goals of the specific review effort. Still, some similarities emerge looking across the 12 review efforts.

The most common non-subject area specific review criteria were: 1) program organization and management; 2) collaboration (with other agencies, family, etc.); 3) student, youth, or client focus/development (life skills oriented, especially with vocational and special education programs); 4) equity and diversity needs; and 5) program evaluation and evidence of success. These standards were often associated with indicators (anywhere from 2-11 per standard) that provided detailed benchmarks on how to meet the standards.

It is also notable that some efforts emphasized replicability as a criteria more than others. The generic screening process and the two review efforts involving secondary and postsecondary institutions emphasized understanding their context as well as any program's context that was being considered as a model for implementation.

An important finding is that these criteria appeared to be most effective when they were based on:

- Research findings or a literature review of the current research (3 review efforts)
- Extensive feedback from practitioners/professionals/leaders in the field (3)
- Provisions of current federal legislation (2)

Conducting site visits to "promising" programs before making a final decision about their status was a luxury that only three of the wealthier review efforts could afford. An expert review team—which five review efforts took advantage of in some format—often conducted or helped with on-site visits. These expert review teams were always made up of 3-5 people.

Evidence of Success

The evidence of success required to become a promising program or practice varied considerably by type of review effort. A few of the review efforts (the transition and higher education reform programs in particular) placed an extreme emphasis on high quality evaluation data, preferably linked to targeted outcomes. While not every program had a clear project description, evaluation design, and a summary report of evaluation findings, most of the applications tried to ensure that some sort of baseline evaluation standard was met.

One review effort (the high schools that combine challenging academic courses and modern vocational studies) used its own achievement data—based on NAEP—to monitor student progress in membership schools. Another application required a separate narrative explaining the evidence of the program's effectiveness. But most frequently, review efforts required applicants to indicate how they met the standards and indicators used as effectiveness criteria.

Incentives for Programs

Some of the review efforts were better than others at providing incentives for programs or practices to apply for promising or exemplary status. Programs were apt to pose the following questions when considering application: What do we get after we are reviewed? Do we get to share resources with other selected programs? How often do the selected programs talk to each other?

Many of the review efforts disseminated information to programs in the following ways:
- At least five efforts had *conferences*, and one effort (the advanced technology centers) required participation in a conference at least once a year.
- For many review efforts, *newsletters* were the most popular vehicle for publicity and sharing between programs.
- At least five of the review efforts set up *program profiles*, in either a report, book, or database format (see Appendix A).
- At least two efforts established home pages on the *Web* and used the site to share information, answer questions, etc.
- One review effort sent out a *press release* of its "exemplary" programs to over 200 professional newsletters and the selected institution for publicity purposes.

Membership programs (there are three such review efforts) were more likely to receive some of these dissemination benefits because they frequently pay a membership fee (between $500-$1,000 for two of the review efforts). They were also more likely to participate in self-assessment/professional development activities. Two review efforts (the early childhood programs and youth development/ employment practices) required or strongly suggested the completion of a self-assessment before completing an application.

Lessons Learned for "Promising" and "Exemplary" Standards and Expert Panels

- There was little differentiation or discussion regarding the differences between "promising" or "exemplary" programs or practices in the review efforts studies. Only two of the review efforts used the term "exemplary" to describe the programs they selected.

- It is most effective to base program evaluation criteria on a careful examination of the research base, ideas from knowledgeable practitioners or leaders in the field, and current federal legislation provisions, whenever possible.

- A rating system (only three reviews had one) and/or the use of experts to conduct the reviews (five programs used in some capacity) made review efforts more credible.

- A good application process required a program to complete a self-assessment as part of the application process, which led to program benefits before the application was reviewed. Two of the applications required or strongly suggested a self-study as a component of their application materials.

- Three other less popular processes that appeared to be effective were: 1) the requirement of sponsorship by a pre-existing member or promising program; 2) the use of achievement data to back up results; and 3) the use of site visits (when costs allowed) to verify information in the application.

Background Information on the Review Activity

Title: National Consortium for Product Quality (NCPQ)
Funder: Office of Vocational and Adult Education, U.S. Department of Education
Implementer: Center on Education and Work, University of Wisconsin-Madison. Barbara Dougherty (bdougherty@mail.soemadison.wisc.edu), Project Director, 1-800-446-0399, http://www.cew.wisc.edu
Purposes of Review Process: To provide practitioners nationwide with benchmarks for assessing their school-to-work curriculum, and to enhance the design of vocational curriculum material through infusing curriculum standards emphasizing content, instruction, student assessment, and equity and diversity.

What Type of Resource is to be Reviewed?

The National Consortium for Product Quality (NCPQ) is interested in reviewing high-quality curriculum for the field of School-to-Work education. The NCPQ requests anyone who is aware of School-to-Work curricular materials worthy of national attention to send them to the NCPQ for review. The instructional products sent to the NCPQ undergo a two-stage review (see below for more details).

What Criteria are Used?

The National Task Force of the NCPQ developed the list of standards and indicators currently used in the review process. These standards statements are broad, qualitative ideals of what is valued in instructional products. Each standard has several indicators (anywhere from two to eleven, in question form) representing essential attributes that support the standards. The indicators can be used to help someone evaluate their materials in an objective or measurable manner.

The four standards are the Content Standard, Instructional Standard, Student Assessment Standard, and the Equity and Diversity Standard. Their four standard statements are:

> *Content Standard*—School-to-Work curricula must focus on the integration of academic foundations into career development, life skills, and occupation competencies.
> *Instructional Standard*—School-to-Work curricula, through active and applied learning experiences in school, community, and work-based settings, must enable students to acquire problem-solving, communication, and reasoning strategies.
> *Student Assessment Standard*—Assessments within School-to-Work curricula must be student-focused in measuring attitudes, knowledge, and skills, as well as their application to problem solving within the classroom and workplace environment.

Equity and Diversity Standard—School-to-Work curricula must reflect content which portrays and celebrates the active participation of all individuals in the nation's workforce, communities, and educational institutions.

The Equity and Diversity Standard addresses the use and effectiveness of the material with diverse users or in diverse contexts. This standard has two indicators: 1) to what extent is the material balanced to reflect the experiences, contributions, voices, and perspectives of all groups?, and 2) to what extent does the content challenge traditional cultural assumptions?

The aforementioned criteria are published in a "Standard Times" newsletter and included in the Submittal Application Form.

How are Decisions Made?

As briefly mentioned above, materials sent to the NCPQ undergo a two-stage review. Phase 1 is conducted by the National Consortium staff, and includes a preliminary review of all products using the Standards and Indicators formulated by the National Task Force. This phase of the evaluation provides a general indication as to whether or not the instructional product reflects the quality standards.

The nominator of each product submitted receives a completed Phase I Review Feedback Form. The feedback is helpful in a number of ways. First, it may be helpful in considering the curriculum for adoption, or in making revisions and enhancements to the submitted curriculum. It also can guide future curriculum development efforts designed to expand or supplement the initial curriculum.

If materials receive high scores in the Phase I review, then they are passed on to the National Consortium's Panel of Reviewers. This Phase II review consists of an in-depth assessment of the product by three to five experts. Comprehensive Product Profiles are then prepared and disseminated nationally for products emerging from the Phase II review. The "Product Profiles" provide instructors, administrators, curriculum specialists, etc. with detailed information on the product's content, instructional design features, format, and availability.

Reviewers, Recognition, and Dissemination

Reviewers. The NCPQ Reviewers Board is made up of a nationwide network of education, business, and labor professionals. The reviewers help evaluate the instructional materials on an annual basis. They lend their particular expertise to the different materials sent in by all of the programs. Selection as a reviewer for certain materials is based on familiarity with both the content and instructional design of that product. The NCPQ is interested in expanding this network, and encourages all those qualified to apply to become a reviewer.

Post Review Efforts. Those materials which receive favorable reviews are included in product profiles which are disseminated nationally (see below for more details). The profile includes information on the content, instructional design features, format, and availability of the product.

Total Review Effort. The product reviews are disseminated nationally through the National Center for Research in Vocational Education (NCRVE), the six National Curriculum Centers, and electronically via VocNet.

Unique Features

- Two-phase review effort with initial screening before expert review
- National dissemination of product profiles

Background Information on the Review Activity

Title: Promising and Effective Practices Network (PEPNet)
Funder: U.S. Department of Labor, Employment and Training Administration, The Ford Foundation, The John D. and Catherine T. MacArthur Foundation, The AT&T Foundation, The Clark Foundation, and The Prudential Foundation
Implementer: National Youth Employment Coalition (NYEC), which is a network of over 80 youth employment/development practitioners, researchers, and policy-makers from around the nation dedicated to promoting policies and initiatives which help youth succeed in becoming lifelong learners, productive workers, and self-sufficient citizens, Edward DeJesus, Director, PEPNet, Alan Zuckerman, Executive Director, Kate O'Sullivan, Policy Associate, (202) 659-1064
Purposes of Review Process: To articulate the principles and practices that make an effective youth development/employment initiative; to offer a tool to organizations to analyze and evaluate their practices against those identified principles and practices; to identify and recognize effective youth programs; to strengthen the network of youth initiatives; and to provide information about effective initiatives and practices so that organizations can benchmark, replicate, and learn from each other.
*Impetus for Review Process: The development of PEPNet began in August, 1995 with a retreat that focused on developing criteria of effective youth employment/development practice. NYEC was responding to a policy environment coming to terms with a wave of negative studies of JTPA and other youth programs, and thus operating in a context of high need and a policy debate questioning the effectiveness of programs (i.e., there is a prevailing misconception that "nothing works" in preparing in- and out-of-school youth for jobs, careers, and self-sufficiency).

What Type of Resource is to be Reviewed?

Any organization involved in youth employment/development serving youth aged 14-25 is eligible to apply to PEPNet.

What Criteria are Used?

The National Youth Employment Coalition (NYEC) convened a PEPNet Working Group of NYEC members and others to identify the criteria of effective youth employment activities. They worked long and hard to distill the breadth of thinking and research they heard and knew of, in order to identify the criteria of effective initiatives and develop rigorous measures of practice.

The four broad PEPNet quality categories that research and field practice suggested all youth initiatives should strive to obtain are Organization and Management, Youth Development, Workforce Development, and Evidence of Success. The four categories are described as:

Organization and Management—Successful initiatives have engaged leadership, qualified and committed staff, work in collaboration with others, and use information and data to continuously improve;
Youth Development—Successful initiatives are characterized by conscious and professional reliance on youth development principles to identify or shape program activities and to drive the kinds of outcomes that are sought for young participants;
Workforce Development—Successful initiatives consciously prepare youth for the workforce, connect successfully to employers, make appropriate use of training, workplace exposure, work experiences, education, and consistently emphasize the connection between learning and work; and
Evidence of Success—Successful initiatives collect and make use of credible data or other measure that reflect the soundness of their goals, their operational effectiveness, and their ability to achieve desired outcomes.

Each category has several criteria (from five to eleven) that require further discussion or analysis for either the application or self assessment. For example, a few of the questions under the Youth Development quality standard are 1) describe program activities that promote cultural/ethnic awareness, 2) how does the program involve family and peers?, and 3) how are young people recruited for the initiative?

Cutting across the categories, the Working Group felt, was the importance of "goal congruence:" a clear statement of plausible aims, and a logical connection between those aims and the services and activities that made up the initiative. The Working Group identified two of these categories with special significance. First, youth development is a critical dimension and needs to be an integral and practical part of any effective initiative. Second, PEPNet needs to take "evidence of success" seriously and stress the importance of success measures—to programs, funders, and policy-makers.

In some cases criteria could be found that were field-tested or widely accepted. In the workforce development area, the SCANS competencies were a known and tested quantity, and their use was a strong indicator of the strength of workforce development. Similarly, the approach to organization and management drew on principles of total quality management.

The criteria are published in the Application and Self Assessment booklet (refined each year) and mentioned in the PEPNet '96 "Presentation of 18 Effective Initiatives: Lessons Learned" booklet, which also outlines common features of the '96 Recognized Initiatives.

How are Decisions Made?

Organizations must fill out a self-nominating application and be willing to participate (if chosen for public recognition) in the learning process for others. The second part of the application package. the Self-Assessment, is to be utilized by any organization (independent of the application process) for the purpose of self-assessment/self improvement. It is suggested that organizations first start the self-assessment process before determining if they want to submit an application form.

After the application is submitted, it will undergo an initial screen to ensure that the information is complete and that the activity is eligible. Then each application will be reviewed by a panel of youth employment/development practitioners. researchers, and policy-makers who know and understand youth programming. Reviewer comments will be shared with the applicant. A national recognition event (awardees announced September 1) will be planned by NYEC.

PEPNet was widely advertised. More than 10,000 youth organizations were notified about the process and the availability of the PEPNet Self Assessment and Application. More than 800 requested the materials in 1996 and 1997. Nearly 100 programs have formally applied to PEPNet in the past two years. Each application was rated by a team of three youth professionals, using the screening criteria, and in some cases applications received a second screening. That process led to the selection of 18 exemplary PEPNet programs in 1996. The 1997 Review Process is currently underway.

Reviewers, Recognition, and Dissemination

Reviewers. The PEPNet Working Group is made up of a broad cross-section of youth experts, representing youth employment and youth development practitioners, business, education, policy analysts, and researchers from across the U.S. Attempts are made to diversify each review team of three—an ideal team consists of a practitioner, policy-maker, and a business/government employee. Each reviewer must take part in a one-day training session. In 1997, sessions were held in San Francisco and Washington, DC.

Post Review Efforts. In September, PEPNet recognizes initiatives that meet its standards. These initiatives will be the nucleus of a youth employment quality movement. They will illustrate effective practice and provide the training ground for youth employment administrators, policy-makers, and workers. PEPNet information will be shared through papers, the Internet, round tables, workshops, and site visits.

Total Review Effort. PEPNet will soon develop a Network to provide support and guidance for other youth programs seeking to strengthen operations, service offerings, and results. All PEPNet applicants and other interested parties will be invited to stay engaged in PEPNet, receive written materials, and participate in workshops, field visits, and other developmental events. The annual PEPNet Recognition process will continue its development, revising the Application and inviting a broader range of programs to apply for recognition and join the youth employment quality movement.

Unique Features

- PEPNet has a profile booklet they disseminate nationally (included in Appendix A)
- Working Group developed criteria based on standards that research and field practice suggested
- A Self-Assessment is included with the Application package (which is included in Appendix B)
- Formal one-day training for reviewers in San Francisco and Washington, DC

Background Information on the Review Activity

Title: High Schools That Work (HSTW)
Funder: The Southern Regional Education Board (SREB)-State Vocational Education Consortium is a partnership of states, school systems, and school sites, and is supported by the DeWitt Wallace-Reader's Digest Fund (have provided a six-year commitment thus far).
Implementer: The SREB is the nation's first interstate education compact. Created in 1948 at the request of Southern governors, SREB helps education and government leaders work cooperatively to advance education and improve the social and economic life of the region. Gene Bottoms, Director, (404) 875-9211.
Purposes of Review Process: The SREB assists state leaders by identifying and directing attention to key issues; collecting, compiling, and analyzing comparable data; and initiating studies and discussions that lead to recommendations for state and institutional action.

What Type of Resource is to be Reviewed?

The SREB High Schools That Work (HSTW) program is the "nation's first large-scale effort to combine challenging academic courses and modern vocational studies to raise the achievement of career-bound high school students." The HSTW program has involved more than 500 high schools in 21 states to this date. It was started in 1987 with 28 sites in 13 states.

What Criteria are Used?

The two major goals of the HSTW program are: 1) to increase the mathematics, science, communication, problem-solving, and technical achievement and the application of learning for career-bound students to the national average of all students, and 2) to blend the essential content of traditional college preparatory studies—mathematics, science, and language arts—with quality vocational and technical studies by creating conditions that support school leaders, teachers, and counselors in carrying out the key practices.

The "centerpiece" of the HSTW program is a curriculum that blends essential college preparatory content in mathematics, science, and language arts with modern vocational studies. The curriculum requires at least four college preparatory English credits; at least three credits each in mathematics and science, with two credits in each subject from college preparatory courses; at least four credits in a vocational major; and at least two credits in related vocational or technical fields. The curriculum is supposed to be based on a functional, "hands-on" teaching methodology. Students are also required to major in either a vocational of an academic field of study, and the general education track is eliminated.

The requirements for states to participate in the HSTW program are stringent. States must agree to establish a state-level council to approve school sites, review annual progress of schools in the state network, and coordinate technical assistance; put a representative on HSTW's board of directors; conduct on-site reviews of schools in the state network; convene an annual meeting of school leaders and teachers to discuss "best practices"; participate in all regional assessment programs and help schools use assessment data; and help schools work through state and local policy barriers.

Participating school sites must also agree to the following stipulations:

Examine goals and key practices—site leaders must decide that the program is viable for the school and community and then commit to a 5-year effort to install key practices and eliminate the general education track;
School board support—school board must support academic and vocational teachers with staff development, materials, and time to implement key practices;
Planning and implementation committee—the site must organize this committee and include representatives of business and industry and postsecondary education (subcommittees must address curriculum, guidance, evaluation, and staff development);
Action plan and staff development plan—the committee prepares these to help teachers carry out the key practices;
Participation in assessments—help determine baseline data and a system to measure progress in raising student achievement;
Active network member—site must agree to share information and ideas within a state and multi-state network; and
Access to modern vocational courses—sites must work closely with employers and two-year postsecondary institutions to provide these courses either at the high school, the vocational center, a postsecondary institution, or in job apprenticeship training.

How are Decisions Made?

Individual states select their own sites and have their own unique application processes. SREB collects data from a number of sources to monitor the achievement of career-bound students at HSTW sites:

1. *The HSTW Student Assessment*—given to HSTW vocational program students to measure achievement in mathematics, science, and reading (based on NAEP and prepared by the Educational Testing Service);
2. *Student reporting*—what they were taught, how they were taught, etc.;
3. *Transcript analysis*—provides information on coursework and relates courses taken to achievement scores;
4. *Administrator, teacher, and counselor opinion*—on integrating academic and vocational education, time devoted to teaching basic academic skills, and staff development needed;

5. *Graduate reporting*—one year after students graduate they report what they are doing and how well they were prepared by their schools; and
6. *On-site reviews*—conducted to determine how key practices are being implemented, identify concerns, and recommend solutions.

Schools are held accountable for the achievement data of their students. For instance, a report on improving career-bound students' learning in mathematics published by HSTW [Reference this report here!] found that, in 1994, SREB's 197 new sites (approximately 12,000 students participated) performed significantly better than vocational students in the national sample. SREB also determined that only 11% of the new sites met the lofty SREB mathematics goal.

The National Center for Research in Vocational Education (NCRVE) designs the evaluation for HSTW, analyzes data, prepares reports, and provides technical assistance and materials to states in the program.

Reviewers, Recognition, and Dissemination

Reviewers. Both SREB staff and state-level people (usually to provide technical assistance) visit HSTW sites frequently, and they often visit schools together in teams. The HSTW staff is made up of 16 people so only a certain number of schools can be visited by those staff in any one year. As mentioned before, individual states each have their own review process.

Post Review Efforts. Membership in the HSTW network improves schools by allowing them to share successful ideas and approaches, and helping them with the incremental planning process—planning, doing, reviewing, making new plans, and revising old ones. Membership also allows schools to review many research reports and site development guides that describe school efforts and successful practices.

SREB specifically provides state, school system, and school site partners with "leadership, guidance, information, and assistance." Services include a framework of proven strategies and a recommended curriculum (see above); assistance in developing a site action plan and site-focused staff development plan; technical assistance in curriculum design, team building, and evaluation; an evaluation process that measures student achievement; and meetings, workshops, and conferences designed to help sites find solutions to common problems.

Total Review Effort. SREB uses the achievement data results (and the other evaluative criteria listed above) to monitor the progress of the HSTW individual sites and program as a whole. For instance, they note in their materials that between 1990 and 1993, the seven most-improved HSTW schools closed the gap between their career-bound students and the HSTW goal by 65% in reading, 36% in mathematics, and 70% in science. Evaluative efforts like these help member schools understand that they are part of an ongoing network of "reforming" schools.

Unique Features

- SREB has a HSTW Web page at http://www.peach.net/sreb/hstw/high.html
- Use of student achievement data, and not just NAEP, but their own test/exam data, to show results (definitely most convincing evaluative data of all the review efforts)
- Hired NCRVE as evaluator to look closely at the projects
- An annual conference that focuses on staff development

Background Information on the Review Activity

Title: A Taxonomy for Transition Programming: Linking Research and Practice
Funder: Transition Research Institute
Implementer: The Transition Research Institute (TRI) is housed at the University of Illinois. Paula Kohler, Research Assistant Professor, (217) 333-2325
Purposes of Review Process: To identify effective transition strategies and test the usefulness of the Taxonomy by identifying particular strategies associated with each practice and the assessment of their effectiveness within specific contexts.
*Impetus for Review Process: There has been a lack of evidence to support "best practices" in transition services for individuals with disabilities. Kohler conducted a four-phase study that led to the Taxonomy for Transition Programming in response to this concern. The Taxonomy is a "conceptual framework of transition practices generated and evaluated by transition professionals and researchers." [p. 64]

What Type of Resource is to be Reviewed?

Practitioners, administrators, researchers, and policymakers submitted various types of transition projects or programs (strategies or services) for students with disabilities. Forty-four projects submitted information in response to a request for "nominations."

What Criteria are Used?

As mentioned above, Kohler describes her Taxonomy as a conceptual framework of transition practices that begins to determine "best practices." The Taxonomy consists of five categories and their respective practice statements:

> *Student Development*—broken down into life skills instruction, employment skills instruction, career and vocational curricula, structured work experience, vocational assessment, and accommodations and support;
> *Student-Focused Planning*—IEP development, student participation, and accommodations and planning strategies;
> *Interagency Collaboration*—individual-level planning, inter-organizational framework, collaborative service delivery, organization-level planning, and human resource development;
> *Family Involvement*—family training, family involvement, and family empowerment; and
> *Program Structure and Attributes*—program philosophy, program policy, strategic planning, program evaluation, resource allocation, and human resource development.

Each of the practices listed above in the five categories also includes a substantial list of subpractices. For instance, the "Resource Allocation" practice under "Program Structure and Attributes" includes the following subpractices: creative use of resources, multiple utilization of funds, sufficient allocation of resources. student/family role in resource allocation, and resources transferred from sheltered and or segregated facilities to community-based and/or integrated settings.

The TRI conducted a review because they were searching for answers to five questions:

1. Which of the practices in the Taxonomy are actual projects or programs implementing?
2. How are the programs implementing specific practices?
3. What are the targeted outcomes?
4. What methods are used to evaluate effectiveness?
5. With respect to the targeted outcomes, what are the evaluation findings?

[Taken directly from p. 74]

Kohler and associates created the "Effective Transition Practices Nomination Form" (See Appendix B) in order to answer the above questions. The first portion of the application asked about demographic information, but the most significant portion required a response to the five-page Taxonomy, which was included. Respondents were asked to indicate which practices in each of the five transition practice categories they were implementing, and describe how they were being implemented. The last few questions focused on evaluation strategies, target outcomes, and evaluation findings.

How are Decisions Made?

Kohler and associates created an evaluation matrix which allowed them to summarize information about each project and aggregate the data across projects. They identified the projects that provided detailed information relevant to the five questions. They looked specifically for projects that had "articulated evaluation components and strategies and that also provided details about practices, strategies, and targeted outcomes." [p. 75]

As stated above, 44 projects submitted applications. Only 18 of the 44 included information on evaluation strategies, and that information varied tremendously in scope and focus. Eight projects described thorough and comprehensive evaluation strategies which evaluated their program's effectiveness. The descriptions usually included a clear project description, an evaluation design, and a summary of the evaluation findings related to targeted outcomes.

Summaries or profiles of the eight projects are included in the report and include names and contact information, project purpose and program overview, unique program components, practices that are included in the Taxonomy, targeted outcomes, and a description of evaluation methods and findings. One of the eight profiles is included in Appendix A.

Reviewers, Recognition, and Dissemination

Reviewers. Kohler and two of her colleagues at the University of Illinois were the reviewers.

Post Review Efforts. The eight projects included in the report "contribute substantially to the pool of innovative and diversified transition-related programs and can be used as important resources for those implementing transition programs." [p. 76] The names and contact information for all 44 submissions are also listed.

Findings are published in a monograph that is disseminated via ERIC and a Web site (see below for address). The Web site includes the full text of the Taxonomy, a description of the review process, and project descriptions. 300-400 copies of the monograph are also disseminated to contacts in the field.

Total Review Effort. This effort includes a big push for program evaluation ("Only by evaluating what we do can we determine what works best.") [p. 76] A new federal project, the National Transition Alliance (see Web address below), will conduct a new search for exemplary School-to-Work programs.

Unique Features

- One of the eight project descriptions is included in Appendix A
- The application/nomination form (see above) is included in Appendix B
- The "MCITT Transition Self-Assessment Instrument" developed directly from the Taxonomy (Ch. 2) is a transition services assessment process
- The Web site for the TRI and the Taxonomy is http://www.ed.uiuc.edu/coe/sped/tri/transindex.html
- The Web site for the National Transition Alliance is http://www.dssc.org/nta/

Background Information on the Review Activity

Title: National Coalition of Advanced Technology Centers (NCATC)
Funder: The Center for Occupational Research and Development (CORD), which is a nonprofit public-service organization dedicated to excellence in education and training for highly skilled workers through new and integrated curriculum materials and processes.
Implementer: NCATC is a division of CORD and was established in 1988. NCATC is a coalition of postsecondary institutions, nonprofit organizations, and corporate entities that seek to promote the infusion of proven advanced technology in American industries through community, technical, and junior colleges, and universities. Rob Auld, Executive Director, NCATC, 1-800-972-2766 or (817) 772-8756; Hugh Rogers, Membership Director, (352) 237-2111, x223; http://www.cord.org/ncatc.html
Purposes of Review Process: To determine which applications/programs will become members of NCATC. NCATC functions as a network of mutual support to help advanced technology centers (ATCs) achieve their fullest potential as mechanisms for technology application, education, and training.

What Type of Resource is to be Reviewed?

Applications to join NCATC are reviewed two to three times a year (see below for more details on application process). NCATC currently (as of 2/10/97) has 108 members in at least 33 states, and has set a goal of 200 members by its 10-year anniversary (1998). Applicants can apply for one of four different membership categories: 1) *affiliate*—international institutions and non-profit entities. 2) *associate*—educational institutions that may be developing ATCs, 3) *corporate*—companies interested in collaborating with ATCs, and 4) *full*—four-year institutions and community, technical, and junior colleges with existing ATCs. NCATC dues for Associate Members are $500 per year.

What Criteria are Used?

The application for center membership is fairly straightforward. Membership is reserved for "those organizations that fulfill the role of industrial training and service through a major commitment of resources, facilities, and focused efforts." Applicants found to be achieving the Membership Criteria are recommended for Associate Member status. Institutions seeking full member status should be sponsored by an existing full member institution (see fourth criteria below).

The Membership Criteria are as follows:

Evidence of ATC Role—commitment to the philosophy of industrial training and service, commitment of the governing and administrative structures of the institution, commitment to ATC development by the governing and administrative structures of the institution, and the commitment of resources for ATC development;

Plans for Development of an ATC—outline plans and timelines for development of an ATC at your institution;
Response to Regional Industrial Needs—describe major industrial needs of the region and how proposed ATC programs and services will address those needs; and
Sponsoring Institution (if appropriate)—provide a reference for an individual at an NCATC member institution or at another organization who could help us learn more about your institution.

Applicants are instructed to list specific indicators of compliance with each criterion. They are also encouraged to append supporting materials (e.g., flyers, brochures, newspaper articles) or other information regarding their Center that might help the Membership Committee (see "Reviewers" below).

How are Decisions Made?

Applications are reviewed on a case-by-case basis. All applications are sent to Erin Mathis, Resource and Communications Coordinator, who conducts an initial screening. She then forwards the applications that meet the membership criteria to Hugh Rogers, Membership Director.

Packets of information on promising applications are put together and sent to Board members before the meetings of the Executive Council. After some discussion on each application at the Council meeting, the Board votes whether or not to invite a Center to join NCATC.

Reviewers, Recognition, and Dissemination

Reviewers. The Membership Committee is the Executive Council, which meets approximately two to three times a year. According to Mathis, the majority of their meeting time is spent making membership decisions.

Post Review Efforts. Membership provides access to planning and development resources required to launch new ATCs and opportunities for other organizations to participate in its workshops, conferences, and activities. NCATC members also have access to CORD's expertise in curriculum development, CORD's school-to-work and workforce development initiatives, CORD's adult academies, CORD's Virtual Teaching Center, and Integrated System for Workforce Education (ISWEC)—a framework for teaching and learning based on national standards in three areas: academic, business, and industry. Because members must share information with other institutions and participate in Coalition activities, they are required to register for at least one of two annual conferences.

Total Review Effort. There is an on-line directory of ATCs including a searchable database of programs and services offered by their parent institutions.

Coalition profiles appear frequently in the NCATC quarterly newsletter. Members also receive resource exchange bulletins and news and analysis of government policies.

Unique Features

- The Coalition has membership dues
- The Coalition has a benchmarking system used to measure ATC performance (the ATC Performance Measuring System)
- Coalition members can also purchase curriculum materials developed by Technological Education Materials Development at reduced prices
- An on-line directory of ATCs is available on their Web page (http://www.cord.org/ncatc.html)
- Requirement of sponsorship by a fellow NCATC member institution in order to reach full member status

Background Information on the Review Activity

Title: Exemplary Career Guidance and Counseling Programs
Funder: Office of Vocational and Adult Education (OVAE), U.S. Department of Education, and National Center for Research in Vocational Education (NCRVE); in cooperation with National Association of State Career Development/Guidance Supervisors (NASCD/GS) and American Vocational Association (AVA)
Implementer: NCRVE Office of Student Services, Carolyn Maddy-Bernstein (maddy2@ux1.cso.uiuc.edu), Director, (217) 333-0807
Purposes of Review Process: To identify, make visible, and recognize exemplary career development/guidance program(s) that are comprehensive and serve all students in their setting; and to use exemplary programs as models for others working to improve their programs.

What Type of Resource is to be Reviewed?

Career guidance and counseling programs that serve individuals in various public or private settings and educational levels (comprehensive high school, specialized vocational high school, secondary level area vocational center, alternative secondary school, postsecondary/technical institute/community college, alternative postsecondary program, adult and continuing education agency) are encouraged to apply. Settings may be a school district or a single school or an institution. Educational levels may include K-adults guidance programs. Programs should be accessible to ALL students in the setting.

The program must also be operational, reflect adequately the requirements and priorities of the State Department of Education in the state in which it operates, and meet the selection criteria described below.

What Criteria are Used?

The following criteria are used in selecting exemplary programs:

1. The program provides measured evidence of effectiveness.
2. The program is cost effective, that is:
   - Measured evidence of effectiveness is greater than the resources being invested;
   - Program costs are in line with other programs with the same or similar outcomes;
   - The program either: (a) operates on revenues available to other districts or institutions; or (b) if receiving special funding, has identified and has committed revenues to continue the program after special funding ends.
3. The program offers students the full range of educational options including vocational-technical careers.

The application (included in Appendix B) consists of four parts. The four sections of the application are (1) general information, (2) demographic information, (3) program information, and (4) components of exemplary career guidance and counseling programs. The information required in each section and points given to each are as follows:

> *General Information (2 points)*—This section asks for basic information about the program (e.g., title of program, program contact), and also requires a letter of support from the state official responsible for career guidance programs at the appropriate level (middle and secondary programs, community college).
> *Demographic Information (8 points)*—This section outlines an objective description of the program including the setting of the program, the population and numbers of students served, staffing requirements, financial information, and outcome data. This section requires a separate narrative of the evidence of the program's effectiveness and supporting documentation (e.g., annual reports, summary of follow-up data).
> *Program Information (4 points)*—This section asks for a written program abstract and a description of the history and operation of the program.
> *Components of Exemplary Career Guidance and Counseling Programs (86 points)*—This section requires detailed descriptions of 16 components of exemplary guidance and counseling programs identified through (a) a literature review of career guidance and counseling programs, (b) research on exemplary schools and exemplary instruction, (c) a review by leaders in the field, and (d) provisions of current federal legislation.

The first five components are listed under the subheading "Career Guidance and Counseling Program Plan," which describes the heart of the program. Three of the components focus on assisting students/clients to achieve career development competencies, including an increase in self-knowledge and self-advocacy, help in educational and occupational exploration, and assistance in career planning, preparation, and transition. The other two components in the subheading are addressing the needs of diverse student populations and program support services.

The second subheading of this portion of the application is entitled "Collaboration, Articulation, and Communication." The four project components here focus on family/parental involvement and support, faculty/staff involvement in career guidance and counseling program, intra- and interagency collaboration, and collaboration with business.

The last subheading, "Institutional Support, Leadership, and Program Evaluation," contains the final seven "exemplary" components. They are institutional support, facilities, financial support, guidance personnel qualifications, professional development (any preservice, inservice, and/or continuing education obtained or conducted by program staff), program evaluation, and follow-up of program completors and noncompletors.

The 16 "Components of Exemplary Programs" have been evolving since 1988, when they focused on exemplary programs for special populations. In 1993, the focus switched to career development, and the components changed accordingly. Maddy-Bernstein reported that they gathered literature and looked at some of the National Demonstration Projects to help determine the components. She also estimated that they have gotten feedback from 200-300 people over the years.

How are Decisions Made?

Each application is evaluated within 8 weeks of submission by three reviewers who are national experts in the field of career counseling. Each reviewer reads approximately 2-3 applications. After all applications have been reviewed, representatives of the highest ranking programs (see description of application process above) are contacted by the search sponsors to set up site visits. (See Appendix C for a copy of the rating form.)

The top ranking programs are visited by a team of reviewers to validate the information in the application and learn more about the program's operation. The programs selected as "exemplary" are notified within 3 weeks of the last site visit, and a press release is sent to over 200 professional newsletters, and to the selected school/institution/college for publicity purposes.

On average, 4-8 "exemplary" programs are selected from the 15-38 applications received yearly. Twenty-five special populations programs have been awarded "exemplary" status between 1989 and 1994. and Maddy-Bernstein estimated that 22 guidance and counseling programs have received the same distinction.

Reviewers, Recognition, and Dissemination

Reviewers. The reviewers are national experts in the field of career counseling. Three national experts or former exemplary program staff review every application, and decide whether or not to recommend a site visit.

Post Review Efforts. Exemplary programs are announced in a variety of newsletters and during national conferences. Program information is highlighted in OVAE's and NCRVE Office of Student Services' (OSS) publications and presentations, and entered into the OSS database of exemplary programs (see below for more details). If a program is not selected and they ask for feedback, they are sent copies of the review sheets.

Within a year of the award a representative of each program will have expenses paid to attend a designated national conference where formal announcement of the award will be made and a plaque presented to the school. The program representative will also present information about the program to conference participants.

Total Review Effort. Information about the exemplary programs (including names and address of contact persons) is disseminated more widely through the database to various individuals, agencies, and other education institutions.

Unique Features

- Includes sample rating form (100 points) in application (application included in Appendix B and rating form in Appendix C)
- Application appears specific enough to help in the development of a good program (has a strong self-assessment component)
- Maddy-Bernstein reported that it was critical to include site visits in the review process (they did so after the first year)

Background Information on the Review Activity

Title: "Screening Promising Practices for Adoption"
Funder: Office of Special Education Programs (OSEP)
Implementer: Judy Smith-Davis, (703) 239-1557
Purposes of Review Process: To provide those involved in efforts to improve the schools through the adoption of new practices and products ("buyers") a number of questions and criteria (consumer standards) that should be applied in their selection process.

What Type of Resource is to be Reviewed?

New practices and programs—instructional materials, training programs, workshops, books, audiovisuals, education models, software and other technological packages—distributed by any number of people across the country (these "sellers" may be commercial, professional, for-profit, non-profit, private, federally funded, or otherwise) are the target of this review effort.

What Criteria are Used?

Dr. Smith-Davis argues that buyers should be systematic, objective, and responsible in sorting out the wide range of innovations that are available to them. She believes that three major concerns should drive the selection process:

> *Does It Work?*—What objective evidence can the seller furnish that a practice or product has been demonstrated to be effective in achieving what it claims?
> *What Are the Minimum Standards for Replication?*—Can the seller describe the precise attributes, procedures, and details that represent the integrity of the practice and that comprise its programmatic and management specifications? Has the innovation been prepared specifically for use by others?
> *Does It Fit?*—Can the buyer define the local conditions and contexts and use this information to determine which practices and products will best fit local characteristics? And can the seller provide information on the innovation that will clarify the contexts in which it is most likely to succeed?

Dr. Smith-Davis writes that sellers' practices are not yet ready to be adopted or adapted by others if sellers are unable to respond to the first two questions above. If the sellers respond, then the buyers have something to work with and can determine how well matched an innovation is with the characteristics of the potential user site.

Dr. Smith-Davis developed a "Screening Promising Practices for Adoption" checklist for OSEP during the 1980s on the basis of a review of literature on dissemination/diffusion. The specifics of the checklist are described in the following section.

How are Decisions Made?

The "Screening Promising Practices for Adoption" checklist is to be used by individuals to review products, but not to select any subset of promising or exemplary programs.

The screening sheet is set up to help the buyer: (a) know what you need programmatically; (b) know yourself—your local conditions, contexts, and resources; and (c) ask structural questions whose answers will tell you whether or not a promising practice will fit local characteristics. The sheet lists structural and contextual questions that one should ask about a practice and about oneself, and suggests a scale for rating the relative match between the practice and oneself.

The checklist is divided into ten sections/topics that ask questions of both the seller and the buyer, and asks the buyer to rate the match between the responses. The ten checklist topics are: (1) claims; (2) effectiveness; (3) relative advantage; (4) history of use; (5) minimum standards for replication; (6) compatibility; (7) complexity and convenience; (8) assistance with implementation; (9) costs; and (10) payoff.

An example question from the "complexity and convenience" section asks the seller "how much discomfort and reorganization will occur if it is put into practice?" and asks the buyer "what level of discomfort and reorganization can be tolerated?" All of the checklist criteria are coordinated in such a manner, and require the buyer to rate the match in each of the ten categories.

Reviewers, Recognition, and Dissemination

Reviewers. As mentioned above, the checklist was developed for a diverse group of users/individuals and not a specific review team attempting to identify the "best" practices or programs.

Unique Features

- The 3-page screening checklist is included in Appendix C of this report.
- This is the only review effort that was fairly generic in scope (i.e., it was not limited to one specific content area or topic).

Background Information on the Review Activity

Title: Special Education Exemplary Programs (December 1995)
Funder: Idaho State Department of Education
Implementer: Special Education Section, Idaho State Department of Education, Jane Brennan, Special Education Specialist, (208) 332-6915
Purposes of Review Process: 1) provide recognition of programs exemplifying the use of sound philosophical and research-based methods in providing special education services; 2) provide model sites and a forum for sharing ideas, information, and resources to encourage other programs to adopt and emulate those identified as exemplary; 3) increase awareness of high quality practices in special education which demonstrate exceptional skill, creativity, and innovative use of resources; and 4) provide Idaho field staff with an opportunity to gain national recognition as quality practitioners and leaders in special education.

What Type of Resource is to be Reviewed?

In order to be eligible for consideration as an exemplary program, applicant programs must have gone through a compliance review by the State Department of Education, Special Education Section, completed any Corrective Action activities required, and therefore be clear of compliance citations. Also, any programs recognized as exemplary in the past three years are not eligible to apply for the current year of recognition.

What Criteria are Used?

The review panel looks for evidence of the program's quality in each of the following areas: student outcomes, least restrictive environment, parent/community involvement, collaboration, instruction, personnel, and resources. The application requires a program to seek one of the following age specific nominations: early childhood, elementary, middle/jr. high school, secondary, transition, or other.

For each of the domains areas mentioned above, the applicant must address the program's strength in that area, the data sources available to support these strengths, and any unique program features. The application provides a comprehensive list of standards, indicators, and examples of data sources for each of the domain areas in order to help the district applicants. This list was compiled by the Special Education Exemplary Programs task group. Exemplary standards for each of the domain areas as well as the relevant indicators (in parentheses) are provided below, but the exhaustive list of data sources is not provided:

> *Student Outcomes*—This program strives for FULL participation of students with disabilities in a wide range of school and community activities (setting clear program expectations and monitoring those expectations on a frequent basis;

including the teaching of skills that will facilitate the full participation of students with disabilities in these activities).
Least Restrictive Environment—This program provides for the successful inclusion of all students with disabilities in some or all of the programs available to students without disabilities (conducting adequate assessments in a variety of settings that include both school and community; emphasizing "with whom" each student with disabilities is educated rather than "where"; providing related services in settings that include students without disabilities).
Parent/Community Involvement—(1) This program provides parent/community education programs that facilitate the learning process of special education students; (2) This program communicates with parents and the community in an effective and positive manner; (3) This program involves parents and the community in the planning, development and evaluation of program improvements for special education students.
Collaboration—This program allows for the smooth correlation of services in an efficient, effective, and professional manner (providing a sufficient number of opportunities for in-district collaboration to occur; providing a sufficient number of opportunities for collaboration to occur between the district and outside agencies).
Instruction—This program's instructional practices allow for necessary flexibility in addressing the individual needs of students (fully accessing the curriculum and curricular materials of the general education programs of the district; using sound practices for planning and assessing student success; employing proven instructional practices and methodologies).
Personnel—(1) This program's teaching staff maintains high but realistic expectations for students with disabilities and themselves; (2) This program's administrative staff provides an atmosphere that fosters both student and staff improvement; (3) This program assists individuals to grow personally and professionally in a supportive environment.
Resources—This program has adequate resources that allow for the effective implementation of student programs (accessing a variety of different funding sources; allocating its resources in an efficient and effective manner).

How are Decisions Made?

The scoring forms used by the on-site review team are included in the application. The total scores will be used to make the final determination of "exemplary" status. Any program chosen for on-site review (based on quality of the application and potential for exemplary status) will receive a copy of its scoring forms from the review team. The scoring system can therefore provide constructive feedback to those applicant programs who were not found to be at a level of effectiveness that could be considered exemplary.

About 20-25 programs are eligible to apply for exemplary status during each yearly cycle. Last year, three programs applied and only one received an on-site review and was eventually deemed exemplary. Two years ago, six programs applied, two received on-site reviews, and one was chosen as exemplary.

The development task group determined some areas to be of more importance than others in deciding program effectiveness and weighted the various core areas accordingly. Points lost in some areas can be gained by implementing unique ideas in others. The application stresses that "uniqueness" is determined by the total review team and not by any single member.

The scoring system is as follows: 40 total potential points for student outcomes; 35 for least restrictive environment; 25 for parent/community involvement; 25 for collaboration; 25 for instruction; 20 for personnel; and 15 for resources. In each domain, 5 of the possible points are a unique features weighting.

Reviewers, Recognition, and Dissemination

Reviewers. Brennan is the facilitator and 4 program directors from around the state (representing different regions) conduct the peer review. Brennan asks the Directors of the different regions for volunteers.

Post Review Efforts. Institutions of Higher Education may access the information (both the application and the selected exemplary programs) and incorporate it into their preservice and inservice training programs. Programs awarded exemplary status receive a plaque at a statewide organizational meeting.

Total Review Effort. Information on exemplary programs is disseminated year-round by Brennan when requested. A flyer/brochure describing each exemplary program(s) is mailed out to every district in the state.

Unique Features

- Application stresses that the list should help clarify what an exemplary program looks like and serve as a guide to assist districts in improving the effectiveness of their programs
- Includes data sources and scoring sheet
- Only application with a References list
- Application also stresses "uniqueness" and rewards it in scoring system
- A Web site is currently under construction (http://www.sde.state.id.us/SpecialEd/)

Background Information on the Review Activity

Title: "Promising Practices and Programs"
Funder: Office for Civil Rights (OCR)
Implementer: Howard Kallem, Supervisory Attorney, Elementary & Secondary Education Policy Development Branch, (202) 205-9641, Sexual and Racial Harassment Team; Sherry Goldbecker, (202) 205-9973, Mathematics and Science Programs
Purposes of Review Process: There were 5-6 review efforts conducted simultaneously in different subject areas. Separate staff worked independently to put together a list of at least 10 promising programs in each subject area. The major purpose of the reviews was to provide regional offices with program examples of effective ways to meet the educational needs of minorities and women.

What Type of Resource is to be Reviewed?

Reviews were conducted to identify promising practices and programs in at least the following areas:

- the prevention or resolution of sexual and racial harassment
- the enhancement/participation of traditionally underrepresented students in mathematics and science
- the enhancement/participation of traditionally underrepresented students in gifted and talented education
- the enhancement/participation of traditionally underrepresented students in student athletics
- desegregation

This was the first time OCR did a review in these different subject areas, and each individual/group used a different process to select the promising practices and programs. Some of the review efforts were conducted by a team of staff while others were conducted by an individual. In order to identify promising practices and programs in mathematics and science, for example, Goldbecker reviewed materials she got from attending the "Quality Education for Minorities" conference and talked to regional offices. She identified programs that might be helpful to school districts in addressing the underrepresentation issue.

A few examples include enrichment classes during the school day, after-school clubs, Saturday academies, a summer residential programs for students and training programs/institutes for teachers. The programs included focus on innovative curriculums, hands-on activities, practical applications of science and mathematics, links with community organizations and businesses, and parent involvement.

What Criteria are Used?

All of the profiles across the different review efforts used a consistent format. The staff working on the different reviews met together and came up with this profile structure. The format for each program description consisted of: (1) Title, (2) Institution/Location, (3) Target Group, (4) Goal(s), (5) Description, (6) Evidence of Success, (7) Limitations/ Caveats (8) Program Contact, and (9) OCR Contact.

In order to identify promising practices and programs in mathematics and science, Goldbecker used a couple of loose criteria—OCR wanted a mix of programs from across the country and a range of ages (from early elementary to postsecondary students). The programs range from relatively small-scale efforts to multi-university consortia and represent many areas of the country. No formal rating system was utilized.

How are Decisions Made?

Decisions were made differently depending on the particular review effort. Most of the decisions were left up to the individuals leading each review effort. In identifying programs, staff were directed not to necessarily identify the "ten best" programs, but to find programs that demonstrated the wide variety of strategies that can be used to address a particular issue.

The "Evidence of Success" section, although mostly anecdotal, was probably the most critical in identifying promising programs. Information was primarily gathered from phone interviews and through written communication. Many staff had regional offices call the regional laboratories to help identify potential programs.

Reviewers, Recognition, and Dissemination

Reviewers. Staff from OCR conducted the reviews. A few of the efforts had a team that worked together to identify the programs.

Unique Features

- Conducted multiple reviews in different subject areas simultaneously
- Most informal of review efforts

Background Information on the Review Activity

Title: "Growing Smart: What's Working for Girls in School" Report (20,000 copies printed in September 1995)
Funder: American Association of University Women (AAUW) Educational Foundation
Implementer: The AAUW Educational Foundation provides funds to advance education, research, and self-development for women and girls, and to foster equity and positive societal change. Priscilla Little, AAUW Research Manager, (202) 728-7616 supervised development of the report; Tanya Hilton, AAUW Director; Sunny Hansen, Joyce Walker, and Barbara Flom, University of Minnesota researchers
Purposes of Review Process: The Foundation wanted a comprehensive overview of the status of girls in public school, and this report is the first comprehensive review of approaches that foster girls' achievement and healthy development. It is the first report of broad national scope to scrutinize the common denominators of approaches that foster girls' engagement in public education.

What Type of Resource is to be Reviewed?

The report is a thematic review of the literature on factors that promote the achievement and healthy development of girls from kindergarten through grade 12. The authors reviewed more than 500 studies and reports written in the U.S. primarily between 1989 and 1994, drawing their findings from both academic papers and reports of community youth development projects. The report presents many exciting ideas that are accessible to school and community leaders, parents and students, and policymakers.

What Criteria are Used?

The authors were broad in their scholarly review and drew on the knowledge of teachers and community youth workers, as well as the studies of researchers. They tapped commissioned reports, expert testimony, program evaluations, and survey information from community youth organizations, in addition to the more traditional review of educational databases that catalog journal articles and published research reports.

Their goal was to find the common ground between the educational literature and the youth development literature. The authors only looked at reports that involved U.S. public schools, had evidence of effective strategies, and concerned K-12 students in order to limit their search. They also agreed to the following set of six criteria:

1. concentrate on literature published primarily between 1989 and 1994;
2. focus on the experience of girls in U.S. public schools;
3. look at the experience of girls in kindergarten through grade 12;
4. maintain the focus on what works for girls rather than what works for schools;
5. consider only reports concerned with the positive impact of programs and strategies; and

6. seek examples that emphasize diversity and interconnections of race, class, and gender.

How are Decisions Made?

Three databases were used for computer searches: ERIC (Educational Resources Information Center), PsycLit (psychology literature), and SocioFile (sociology literature). Two members of the five-person research team prioritized titles and abstracts using the criteria listed above. High-priority documents were placed in a file for further review and were read in their entirety.

The authors supplemented the computer search materials with additional readings by contacting national youth development organizations, government agencies, foundations, minority resource centers, centers for school change, and gender equity offices of all states. They found surveys, commissioned reports, studies, and evaluation projects relevant to the study.

Upon review of the 500 academic studies and youth development project reports, they filed each summary under one of five dimensions: people, places, programs, policies, or priorities. Documents and books basically came in one of five forms:

> *Summary literature reviews*—which provided conclusions from previous research on girls and on school climate.
> *Empirical studies*—which displayed quantitative information about girls' achievement and attitudes.
> *Qualitative works*—narrative pieces which offered descriptions of girls in particular school settings.
> *Conceptual and theoretical articles*—which discussed relevant issues and themes written by experts in the field of education, psychology, sociology, and youth development.
> *Project reports*—which offered concrete examples of what works for girls.

Discussions by team members who reviewed the documents in each category and shared all major ideas found in individual files led to the creation of five major themes found in successful programs and practices: 1) celebrate girls' strong identity, 2) respect girls as central players, 3) connect girls to caring adults, 4) ensure girls' participation and success, and 5) empower girls to realize their dreams.

A selected list of programs that embody the principles in the report are presented in one chapter. The report states that the list is not comprehensive (although the review described above was fairly extensive) nor intended to be an endorsement of any programs.

Reviewers, Recognition, and Dissemination

Reviewers. The reviewers were the researchers from the University of Minnesota and the AAUW Educational Foundation. The first draft of the study was presented to five focus groups, three made up of girls and two of women. Eight diverse and expert peer reviewers also reviewed the final manuscript.

Post Review Efforts. The report was very well-received by local and community organizations who especially appreciated the model programs/best practices section. The AAUW Educational Foundation expects to commission papers from a roundtable of single sex researchers in November.

Total Review Effort. 20,000 copies of the report were printed and disseminated in order to share both successful principles and exemplary programs.

Unique Features

- Includes a list of "Action Strategies" (for school administrators, teachers, counselors, school staff, political leaders and policymakers, community youth groups, parents, and students)

Background Information on the Review Activity

Title: National Academy of Early Childhood Programs
Funder: National Association for the Education of Young Children (NAEYC)
Implementer: NAEYC is the nation's largest organization of early childhood professionals (membership approximating 100,000 in 1995) and has provided distinguished leadership in the field of child care and early education for more than 60 years, Stephanie Glowacki, Deputy Director, 1-800-424-2460, x305, Pat Mucci, Accreditation Training Coordinator, x315
Purposes of Review Process: Accreditation is a process in which a program's administrators, staff, and parents join with representatives of the Academy to determine whether that program meets nationally recognized criteria for high quality. The benefits of accreditation are: (1) it assists parents in their search for high quality programs for their children; (2) it improves the quality of group programs available for young children and their families; (3) it provides a valuable professional development experience for teachers and directors; (4) it assures contributors to early childhood programs of a sound investment; and (5) it provides professional and public recognition for high quality early childhood programs.

What Type of Resource is to be Reviewed?

All types of early childhood programs—child care centers, preschools, kindergartens, and before- and after-school programs—are eligible to apply for accreditation. Programs may apply whether they operate on a full- or part-day basis or whether they are for profit or nonprofit. Programs must also be operational for at least 1 year before accreditation can be granted.

What Criteria are Used?

The Academy defines a high quality early childhood program as one that meets the needs of and promotes the physical, social, emotional, and cognitive development of the children and adults (parents, staff, and administrators) who are involved in the program. The specific criteria for these programs address all aspects of an early childhood program: interactions among staff and children; curriculum; staff and parent interactions; administration; staff qualifications and development; staffing patterns; physical environment; health and safety; nutrition and food service; and program evaluation.

Exemplary standards for each of the criteria—except program evaluation—as well as the relevant indicators (in parentheses) are provided below:

> *Staff-Child Interaction*—Interactions between children and staff provide opportunities for children to develop an understanding of self and others and are characterized by warmth, personal respect, individuality, and responsiveness. Staff facilitate interactions among children to provide opportunities for

development of social skills and intellectual growth (conversation, activity, involved children, accessible teachers, affection).

Curriculum—The curriculum, or educational plan, encourages children to be actively involved in the learning process, to experience a variety of activities appropriate to their age and rate of development, and to pursue their own interests in the context of life in the community and the world. Children learn through play that is planned by adults to teach them language, concepts about the physical world, social skills, problem solving, motor coordination, and self-confidence (variety, involvement, child-directed activities, teacher-guided activities, cultural diversity, responsibilities).

Communication with Parents—All communication between programs and families is based on the concept that parents are the principal influence in children's lives. Parents are well-informed about and welcome as observers and contributors to the program (informed parents, home-school communication, welcome access).

Staff Hiring and Qualifications—The quality and competence of the staff are the most important determinants of the quality of an early childhood program and of positive outcomes for children. It is critical, therefore, that the program is staffed by adults who are trained in child development and who recognize and provide for children's needs (careful hiring, trained teachers, in-service training, record keeping).

Staffing Structure—The program is sufficiently staffed and organized to assure that the needs of individual children are met, and to maintain positive interactions and constructive activity among the children and staff (supervision, continuity, small groups).

Program Administration—The quality of the early childhood experience for children is affected by the efficiency and stability of the program's administration. Effective administration includes good communication, positive community relations, fiscal stability, and attention to the needs and working conditions of staff members (written policies and procedures, record keeping, insurance, staff meetings, self-evaluation).

Physical Environment—The indoor and outdoor physical environments should be designed to promote involvement in the daily activities and easy, constructive interactions among adults and children (space, easy movement, activity areas, exercise).

Health and Safety—The health and safety of children and adults are protected and enhanced. Good programs act to prevent illness and accidents, are prepared to deal with emergencies should they occur, and also educate children concerning safe and healthy practices (licensed program, health records, maintenance, supervision, transportation safety, accident reporting, emergency procedures, handwashing, safe equipment, product safety).

Nutrition and Food Service—Children are provided with adequate nutrition and are educated concerning good eating habits (well-balanced meals, social interaction, sound nutritional practices).

How are Decisions Made?

The accreditation process involves three major steps. The first step is a "Self-Study" where program personnel and parents determine how well the program meets the Academy's Criteria. They then make needed improvements and report the compliance with the Criteria on the Program Description form provided by the Academy.

The next step is "Validation" where trained validators make an on-site visit to verify the accuracy of the Program Description. The last step, "Accreditation Decision," involves a 3-person Commission which considers the validated Program Description and makes the accreditation decision based on professional judgment. Programs are not required to demonstrate 100% compliance to be approved.

The accreditation system is a nonprofit, self-supporting, and permanent project. The ongoing operation of the system is sustained by fees paid by participating programs. The cost of accreditation varies depending upon the number of children enrolled in the program, but can reach close to $1,000. An application fee is paid to initiate the process and to receive the self-study materials. A validation fee is paid if based on the results of the self-study a program decides to pursue accreditation.

Reviewers, Recognition, and Dissemination

Reviewers. There are two types of reviewers—Validators and Commissioners. Validators must undergo a full-day training and are most often Directors of Accredited Programs. If not, they must have the following qualifications: (1) at least a BA in early childhood education, (2) experience with young children in group programs, (3) experienced administrator from an accredited program, and (4) flexibility to conduct three site visits per year.

Commissioners are usually experienced Validators and actually make the "accredited" vs. "deferred" decisions. If they are not an experienced Validator, then they must have the following qualifications: (1) diverse background in early childhood education, (2) national perspective, (3) high-level analytical ability, (4) ability to be objective, (5) ability to trust professional judgment of others (Validators), and (6) knowledge of NAEYC criteria.

Post Review Efforts. The programs that gain accreditation are included in the Academy's list of programs and disseminated widely. This list is updated monthly/bi-monthly and is kept in their offices. It is mailed out to anyone who requests it. If a program is deferred, it has one of three options: (1) appeal to another Commission (they review monthly), (2) make improvements and resubmit, or (3) choose to start over.

Total Review Effort. The Academy's accreditation process began operation in 1985. Nearly 5,000 programs have achieved accreditation and an additional 10,000 programs are engaged in the process. More than 900 programs enter the accreditation process each year, with approximately 80% achieving accreditation upon initial consideration. Of the 20% deferred, 90% reapply and are eventually accepted. Half a million children are served each year in accredited programs across the country.

Unique Features

- Accreditation application materials must be purchased from NAEYC
- By far the largest review effort

Background Information on the Review Activity

Title: "Lessons Learned from FIPSE Projects (Volumes I-III)"
Funder: Fund for the Improvement of Postsecondary Education (FIPSE)
Implementer: As part of the U.S. Department of Education, FIPSE provides grants to improve postsecondary education opportunities by focusing on problem areas or improvement approaches in postsecondary education. Dora Marcus, Program Evaluator, (202) 708-5758.
Purposes of Review Process: These volumes describe interesting campus experiments and attempt to draw conclusions by asking what worked, what did not, and why. They are designed to help colleges and universities facing issues similar to those faced by the projects described.

What Type of Resource is to be Reviewed?

All of the FIPSE funded projects are reviewed by FIPSE staff for possible inclusion in the "Lessons Learned" books. The projects are divided into chapters by their reform focus: from Volume II—assessment, college and university teaching, curriculum and teaching in the disciplines, general education, teacher education, and ethics; and from Volume III—school-to-college transition and retention, rewarding effective teaching, improving teaching and learning, improving the undergraduate curriculum, assessment, teacher education, and postgraduate curriculum and instruction.

In Volume I, only 15 projects were selected for inclusion in "Lessons Learned," but in Volumes II and III, 30 and 31 projects were profiled respectively. FIPSE staff do not designate a certain number of profile slots each year to a given subject area, but rather try to select the "best" 30 or so projects without regard to type. For instance, 10 projects focusing on assessment were chosen for Volume II while only 4 were chosen for Volume III. In general, FIPSE draws from several years worth of projects for each volume, and fund about 70 projects per year.

What Criteria are Used?

According to FIPSE's program evaluator, there are two primary criteria for selection in "Lessons Learned." The first criterion is the significance of the reform idea, i.e., what is its potential for other campuses. FIPSE staff are looking for programs that are most ready for adoption and have a wide range of applicability. The second criterion is the quality of the evaluation. The program should have a solid evidence base because FIPSE staff have "stringent rules about good data."

Each profile in the "Lessons Learned" volumes is structured similarly although certain headings may be left out if they are not relevant to a certain reform project. For example, in Volume II, the headings were:

- Purpose
- Innovative Features

- Evaluation
- Project Impact
- Unanticipated Problems
- Project Activities
- Major Insights and Lessons Learned
- Project Continuation
- Recognition
- Available Information

How are Decisions Made?

FIPSE's program officers and evaluator read through project files, final reports, interim reports, site visit reports, and evaluation reports submitted, and write a paragraph summary making a recommendation about inclusion in "Lessons Learned." They also write a thoughtful letter about the final report to the project director. In addition, the rate the project on many variables (project implementation; institutionalization; quality of evaluation and instruction; access gained through project reforms; student perceptions; cost benefits; impact on home campus, on other campuses; project support (financial); actual products and resources resulting from the project; adequacy of FIPSE support to grantees; and a global overall quality project rating. These ratings are recorded on the "FIPSE Final Report Score Card" for each project.

Reviewers, Recognition, and Dissemination

Reviewers. As mentioned above, FIPSE program officers and the evaluator review the projects. This is the first year that program officers are helping the evaluator review the final reports. The evaluator, one program officer, and an outside consultant reviewed them all in previous years.

Post Review Efforts. FIPSE has had to print more copies of each volume as they have come out. For "Lessons Learned III," FIPSE originally printed 16,000 copies, and then another 6,000 in a subsequent printing. There may even be a third printing.

Total Review Effort. All these "Lessons Learned" volumes are on FIPSE's U.S. Department of Education home page (see below for address). FIPSE also has a special Dissemination grant program. There are plans to include a special section of "Lessons Learned IV" on Dissemination projects (maybe covering 4-5 projects) from the above program.

Unique Features

- An example FIPSE profile is included in Appendix A of this report
- FIPSE has developed a draft version of a project "Score Card" that was being pretested when this report was written
- All the volumes are on FIPSE's Web page (http://www.ed.gov/prog_info/FIPSE)

Appendix A: Sample Profiles

BRIDGES...FROM SCHOOL TO WORK

Project Title:
Bridges...from School to Work

Contact Person:
Nancy Carolan, Program Manager

Mailing Address:
Marriott Foundation for People with Disabilities
Department 901.10
Marriott Drive
Washington, DC 20058

Telephone #:
301-380-7771

Fax #:
301-380-8973

Project Purpose:

Bridges...from school to work was developed to attack the staggering unemployment of young people with disabilities by helping them enter work before they exit school and, thereby, establish the foundation for long-term career success. The primary goals of this project are (a) to provide students with disabilities job training and work experience that will enhance their employment potential, and (b) to help employers gain access to a valuable source of employees and learn to work effectively with them.

General Program Overview:

The Marriott Foundation for People with Disabilities was established in 1989 to foster the employment of young people with disabilities. The Foundation operates a transition program, "Bridges...from school to work," that develops paid internships for students with disabilities in their final year of high school. The program is managed in local communities through an administering organization under the direction of the Foundation.

In recent years the project has expanded to include the following sites: Montgomery County, MD; Washington, DC; Chicago, IL; San Francisco, CA; Los Angeles, CA; and San Mateo County, CA.

The Bridges model was designed to collaborate with and complement existing transition efforts. The following three objectives are important to the success of this program (a) to help students gain critical job experience as they prepare to leave school, (b) to help local employers gain access to a valuable source of employees, and (c) to assist employers in making reasonable accommodations for their workers with disabilities. The Foundation operates Bridges under the premise that successful employment for people with disabilities can occur when both potential employers and potential employees are appropriately supported, particularly early in their relationship.

Unique Program Components:

Interests and abilities: Students are placed in positions that match their skills, interest and experience.

- *Integral part of the working staff:* Students are placed in an environment where supervisors and co-workers are an integral part of the placement, training, and support process, and are therefore, comfortable with their presence.

- *Employer and student support:* The student and employer are appropriately supported, especially early in their work experience, to help ensure job success.

Taxonomy Practices Identified:

Student Development

- Life skills instruction
- Employment skills instruction
- Career & vocational curricula
- Structured work experience
- Vocational assessment
- Accommodations & support

Student-Focused Planning

- IEP development
- Student participation
- Accommodations & planning strategies

Interagency Collaboration

- Individual-level planning
- Interorganizational framework
- Collaborative service delivery
- Organization-level planning
- Human resource development

Family Involvement

- Family training
- Family involvement
- Family empowerment

Targeted Outcomes:

- Increased knowledge of how to get and maintain a job
- Increased employability skills
- Increased employment rates
- Increased maintenance of employment
- Increased employer appreciation and awareness of individuals with disabilities
- Increased employee and employer satisfaction
- Increased family involvement

Evaluation Description:

- *Quantitative design:* A wide range of instruments and activities are used to collect data. A comprehensive, computer-based system using customized software is utilized for evaluating all aspects of Bridges' replication and operation. The evaluation system is intended to ensure that (a) project outcome data are collected, analyzed and reported in a timely and accurate manner; (b) all project processes and procedures are thoroughly documented; (c) continual feedback is available to staff from all constituencies regarding project performance; (d) project accomplishments reflect identified goals and objective; (e) targeted action can be taken on a timely basis to address discrepancies between project outcomes and objectives; and (f) Bridges' objectives are evaluated on a monthly and quarterly basis.

- *Qualitative design:* Data are gathered and analyzed on an ongoing basis regarding the reactions of various parties to the project processes and the effectiveness of outcomes. Respondents include (a) youths with disabilities served by the project, (b) their parents or significant advocates, and (c) their employers. The student and employer evaluation forms are used as data collection tools.

Evaluation Findings:

- To date, Bridges has served over 2,100 youth with disabilities, 76% of whom are of minority background.
- Eighty-five percent of the student workers were placed in competitive, unsubsidized internships with over 700 businesses and organizations.
- All placements are at least minimum wage (the project average is $5.20/hour) and usually average 20-25 hours per week of work.
- Four of five youths who completed their internships were offered ongoing employment.

Promising & Effective Practices Network (PEPNet) 1996

Arizona CALL-A-TEEN
649 North 6th Avenue
Phoenix, AZ 85003
Tel: (602) 252-6721
Fax: (602) 252-2952

Mission

The mission of Arizona CALL-A-TEEN Youth Resources is to provide at-risk teens and young adults with comprehensive training and education programs that contribute to the development of long-term economic self-sufficiency. The intent of each CALL-A-TEEN program is to provide a secure and respectful environment that will allow for the development and mastery of educational and work-related skills and the development of citizenship and personal skills that contribute to the commitment to lifelong learning.

Description

The Arizona CALL-A-TEEN initiative consists of two separate program components: JTPA programming (Title IIC, supplemented by an adult Basic Education grant), and education and services provided as a public high school "center of excellence," or CoE, chartered by the Arizona Board of Education. Young people can participate in both components. CoE is designated as a JTPA School-wide Project site by the Service Delivery Area, allowing all learners at the CoE residing in the city of Phoenix to be eligible for JTPA services. Similarly, any JTPA participant who has not graduated from high school can enroll at CoE .

A full range of academic instruction is offered. The emphasis of the school includes effective instructional practices consistent with adolescent development needs, integration of work and learning and essential content-area skills, preparation for the work place, and linkages to address social service, family and health issues.

The JTPA component emphasizes competency attainment in three employability skills areas: Pre-Employment/Work Maturity, Basic Education, and Job Specific Skills. This component also includes a strong case management approach to service provision.

Contact

Bernice Lever
Chief Executive Officer

Community Setting/ Population Served

CALL-A-TEEN participants are enrolled from throughout the greater metropolitan Phoenix area. During the 1994-95 school year 34% of the participants were enrolled but not attending school at the end of the year, 30% were attending alternative schools and 6% entered school as over-aged freshmen. Over half of the students enrolled in programs are between the ages of 16 and 18. The programs serve almost equal numbers of male and female participants.

Staff

Chief Executive Officer
School Principal/Director of Operations
Director of Training and Program Design

Center of Excellence Charter High School (CoE)
Certified Teachers
Employment and Training Specialist
Registrar
Teacher Aides

JTPA
Employment and Training Specialists
Job Developer
Education Specialists
Assessment and Intake Specialists
Clerks

Exemplary Practices

Quality Management

Continuity of Leadership
One of the strengths of the organization is the continuity of leadership. The Chief Executive Officer has been with the agency since 1978. The School Principal/Director of Operations and the Director of Training and Program Design have been with the organization since 1979. Overall, staff providing direct services have been with the agency for an average of six years each.

Youth Development

Family Involvement
Parents and family are encouraged to become a support system for students and participants. Parents/legal guardians are required to participate in the intake process for both components. Expectations regarding behaviors, attendance, peer relationships, outside influences, etc., are discussed. Students' individualized plans require parent/guardian signatures, thus involving the family in the overall education/training plan for the young person.

Workforce Development

Linking School to Work
This initiative structures classes to approximate work environments. SCANS competencies are incorporated within the curriculum for each content area. Credit is awarded for successful community service, volunteer and work experiences based on attainment of specific training/learning activities. The intent is to blur the lines between work and learning, place equal value on academic and vocational instruction and ensure that all students have the opportunity to develop skills that apply beyond classroom walls.

Selected Highlights

Support of Higher Education
In addition to offering a full range of academic instruction, the CoE is supplemented by a Dropout Prevention Program grant funded by the Arizona Department of Education that allows the school to increase supportive services to its students and that provides funding for matriculation at Gateway Community college.

Evidence of Success

Successful Replication
Within the workforce development community and the charter school movement in the state of Arizona the initiative is recognized as an innovative and effective program for young people at-risk. The design of the effort has been adapted in five Service Delivery Areas through out the state.

Positive Reputation within the Community
Possibly the greatest indicator of the initiative's positive relationship within the community is that both programs have waiting lists of over 100 potential participants.

Changes in Participant Attitudes
Based on the results of an exit survey completed anonymously by CoE students at the end of the 1995-1996 school year, 58% of students responded that before starting CoE, they attended school never or sometimes, 28% attended most of the time, and 14% stated that they always attended. After attending CoE, 100% of the respondents stated that they attend school most of the time. When asked if they were satisfied with their school, 80% stated that they were very satisfied or satisfied as compared to 19% who were somewhat satisfied and 2% not satisfied.

Before they started CoE, 28% of the students felt great or good about themselves and 72% felt okay or not so good about themselves. In comparison the exit interview showed that 76% of the students felt great or good about themselves and 23% felt okay about themselves. There were no responses from any student stating that they did not feel good about themselves.

Lessons Learned from FIPSE Projects II - September 1993

City University of New York

The National Project on Computers and College Writing

Purpose

This evaluation and dissemination project aimed to assess the effectiveness of computer-assisted composition programs and to showcase outstanding college programs that employ this instructional strategy. A network of 15 competitively selected institutions of all sizes and types, chosen from among 90 applicants, participated by joining in a centrally designed assessment program and by sharing their computer-use models and instructional strategies with each other and with wider audiences.

Project Activities

Each of the 15 institutions agreed to carry out an assessment activity according to a common plan. Each site identified at least six sections of English composition, three taught using computer-assisted instruction and three by conventional methods. Each section enrolled similar kinds of students within a single institution, but this was not necessarily the case across institutions. A pre-course and post-course test common to all 15 sites was administered to each student, who in addition completed the Descriptive Test of Language Skills Sentence Structure Subtest and questionnaires regarding attitudes toward writing and writing anxiety. Each participating institution gathered site-specific qualitative data. The final sample consisted of 1,700 students, equally distributed between computer-instructed and regular sections.

The essays were evaluated centrally, all 1,700 holistically and 10% using analytic assessment methods. While overall the holistic scores showed no difference between experimental (computer instruction) and control (traditional instruction) groups, the scores did reveal significant advantages of computer instruction for those in developmental classes, adult learners and community college students. The analytic scoring did show significantly better performance by students in the experimental sections. Students in computer-based classes also showed a significant reduction in writing anxiety as compared with those in the control groups.

The dissemination activities of the project were concentrated in a conference, "Computers and College Writing: Curriculum and Assessment for the 1990's," and a monograph, *Computers and College Writing: Selected College Profiles*. The conference enrolled 600 participants from North America and Europe in June, 1990 and was followed by additional conferences in fall, 1990 and spring, 1992. The monograph contains accounts of computer-based composition technology and instructional strategies at 49 institutions in the United States.

Innovative Features

The project is unique in its efforts to assess the effects of computer instruction on student learning and attitudes across a large number of institutions of varied size and type. No base of comparable size exists. The project both addressed questions about the value of computer-based composition instruction and created opportunities for the exchange of information about the way computers are being used in many institutions. Thus it provides an experimental base to justify the use of computers in teaching introductory composition courses as well as furnishing practical resources for teachers and institutions.

Evaluation

The assessment plan and its implementation were not themselves the object of independent external evaluation. Project personnel have, however, been careful to point out the problems inherent in carrying out such a study and employing the kinds of assessment described. Specifically, they point to problems inherent in holistic assessment. While the methodology, when used with trained evaluators, works well in handling a large volume of essays, when used in a pre-test/post-test situation it tends to produce a regression (and advancement) toward the mean pattern. To counteract this effect the project used a number of measures of student growth in addition to the essays.

Holistic scoring, in focusing on overall impressions, fails to identify particular advances in student writing, such as organization, copiousness, or stylistic sophistication. For this reason, 10% of the essays were subjected to analytic assessments, which revealed more significant differences between control and experimental groups. This result led project staff to assess the entire sample analytically, an effort still in process.

Project personnel were well aware of the problems inherent in trying to measure gains over a single semester of instruction, but the logistical problems of maintaining contact with students and administer a common assessment some semesters later were impossible to manage.

Project Impact

The number of participants in the conferences sponsored through the project and subsequent to its completion reflects its far-reaching effects. Project activities and results have also become known through numerous articles written and produced by project personnel.

The project has produced a firmly established network of institutions engaged in computer-based instruction, and a data base of both institutional strategies and assessment results. This information is a valuable resource for colleges and universities to justify purchase of computers for composition instruction and for those seeking information on the technology, models and strategies of such instruction.

Unanticipated Problems

Apart from the difficulties in finding fully satisfactory and manageable assessment mechanisms, which were acknowledged in advance, the project worked much as planned. Integrating the large volume of information gathered, an effort that continues to the present, required considerably more time and energy than expected.

Major Insights

The project has produced convincing evidence of the value of computer-based composition instruction, particularly for less well prepared groups of students.

Project Continuation

In the spring of 1992, the Project sponsored "Computers Across the Curriculum: A Conference on

Technology in the Freshman Year," which addressed computer-based instruction in a range of introductory college courses.

An effort is now underway to use the project's findings as the basis for a program of technology-driven writing instruction in the New York City Public Schools.

Available Information

The project has generated a number of articles, as well as the major monograph, *Computer and College Writing: Selected College Profiles.* General information about the project and copies of the publications may be secured by writing to:

Max Kirsh, Director
National Project on Computers and College Writing
Office of Academic Computing
City University of New York
555 W. 57th Street, 14th Floor
New York, NY 10019
212-541-0320

Each of the project sites has developed curricular materials, from videotapes to course syllabi. These are available from the individual institutions or, in some cases, from the project director, who will gladly supply the list of participants.

-###-

[Austin Peay State University] PREV UP NEXT [Harvard University]

Appendix B: Sample Applications

Transition Research Institute
Effective Transition Practices Nomination Form

Demographic Information

1. Project or Program Title: ________________________________

2. Contact Person: ________________________________

3. Institution or Organization: ________________________________

4. Mailing Address: ________________________________

____

____

____

5. Telephone: ________________________________

6. Please indicate the type of organization through which the practice(s) is implemented.

_______ University, four-year college, or University Affiliated Program
_______ Community college
_______ Education agency (state, local, intermediate, or tribal)
_______ Private not-for-profit agency
_______ State agency
_______ Parent Organization
_______ Other ____

7. Indicate the geographic service delivery area of the program or project implementing the practice(s).

____ Rural area (places of <2,500)
____ Towns and cities of 2,500-50,000
____ Urbanized area (cities and surrounding areas of 50,000-100,000)
____ Metropolitan area (cities and surrounding areas of 100,000+)
____ County
____ Region with a state (i.e., more than one county)
____ State (or outlying area of U.S., e.g., Puerto Rico)
____ More than one state
____ National
____ American Indian/Alaskan Native area (e.g., village, reservation, trust land)

8. Indicate the primary setting(s) in which the targeted transition practice(s) is delivered.

____ Medical clinic
____ Community-based training site
____ Independent living facility
____ Experimental or research laboratory
____ Home-based setting
____ Hospital setting
____ Middle school or junior high
____ High school or other similar secondary educational setting
____ Four-year college or university
____ Trade school proprietary institution
____ Community college(two-year college)
____ Private school
____ Regular education class
____ Residential school or facility
____ Resource room
____ Self-contained class in regular school
____ Special day school
____ Competitive employment workplace
____ Sheltered employment workplace
____ Supported employment workplace
____ Other Setting ________________
____

Project or Program Consumers

9. Indicate the approximate number of individuals with a disability (consumers) served through the project or program during the current year.

________ # of consumers

10. Indicate the approximate percentage of individuals with a disability served during the current year, by gender.

_______ % of male _______ % of female

11. Indicate the approximate percentage of individuals with a disability served during the current year, by their ethnic affiliation.

_____% American Indian/Native American
_____% Asian
_____% Black/African-American
_____% Hispanic
_____% Pacific/Native Hawaiian
_____% White
_____% Multi-ethnic (e.g., Black and Hispanic)
_____% Other ____

12. Indicate the disability category(ies) represented by the individuals participating in the project or program implementing the practice(s).

_____ Deaf
_____ Deaf-blind
_____ Hearing impairment
_____ Mental retardation
_____ Multi-handicapped
_____ Orthopedic impairment
_____ Other health impairment
_____ Severe emotional disturbance
_____ Specific learning disability
_____ Speech impairment
_____ Visual handicap
_____ Autism
_____ Traumatic brain injury
_____ Other ____________________

13. If applicable, indicate the number and type of other individuals to which transition practices have been directed during the current year (e.g., parents, family members, teachers, etc.)

Number Description

_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____
_____ ____

Project or Program Practices

14. Using the taxonomy on pages 5 through 9, please make a check mark (✓) next to the practice(s) for which you are submitting an implementation strategy.

Student Development

LIFE SKILLS INSTRUCTION

- Rights and responsibilities training
- Leisure skills training
- Social skills training
- Self-determination skills training, including goal setting and decision making
- Self-advocacy skills training
- Community-based independent living skills training
- "Understanding your disability" training
- Learning strategies skills training
- Student training to use natural supports
- Mobility training

ACCOMMODATIONS AND SUPPORT

- Development of environmental adaptations
- Provision of assistive technology devices
- Identification and development of accommodations
- Identification and development of natural supports for all transition outcome areas
- Transportation services
- Infusion of related services into career and vocational development (e.g., OT, PT, speech therapy)
- Peer mentorships
- Use of mentors

EMPLOYMENT SKILLS INSTRUCTION

- Work-related behaviors training
- Job seeking skills training
- Work attitude and work ethics training
- Employability skills training
- Community-based vocational skills training
- Longitudinal vocational training
- Vocational skill training

VOCATIONAL ASSESSMENT

- Vocational assessment portfolios
- Situational assessment
- Ongoing assessment
- Continuous assessment of employment opportunities and job requirements
- Curriculum-based vocational assessment
- Assessment for assistive technology devices

CAREER AND VOCATIONAL CURRICULA

- Community-referenced curricula
- Vocational training begins by middle school level
- Career education curriculum
- Career and vocational curricula infused throughout academic subject areas
- Tech prep curriculum options
- Longitudinal career education
- Participation in mainstream vocational class or program
- Cooperative education

STRUCTURED WORK EXPERIENCE

- Apprenticeships
- Paid work experience
- Multiple, varied community work experiences
- Work study program
- Job placement prior to school exit
- Job shadowing
- Job placement services
- Job matching

Student-Focused Planning

IEP DEVELOPMENT

- Transition-related goals and objectives specified in the IEP
- Post secondary education or training goals and objectives specified in the IEP
- Community participation goals and objectives specified
- Vocational goals and objectives specified
- Residential goals and objectives specified
- Recreation and leisure goals and objectives specified
- Post secondary options identified for each outcome area
- Educational experiences correspond to transition-related goals
- Transition goals are measurable
- Financial issues addressed in planning
- Medical needs addressed in planning
- Guardianship addressed in planning
- Specified goals and objectives result from consumer choices
- Progress toward or attainment of goals is reviewed annually
- Responsibility of participants or agencies specified in the planning document
- IEP supported by individual career plan

STUDENT PARTICIPATION

- Self-determination facilitated within the planning process
- Planning decisions driven by student and family
- Planning process is student-centered
- Planning process is student-directed
- Student participation in planning
- Student involvement in decision making
- Documentation of student interests
- Documentation of student preferences
- Student made aware of post secondary educational institutions and services available
- Preplanning activities for students
- Career counseling services provided to student
- Student self-assessment of preferences
- Student self-assessment of interests
- Student self-evaluation of his or her progress
- Student prepared to participate in planning via curricular activities (e.g., communication, interactive skills, etc.)
- Identification of student interests

ACCOMMODATIONS AND PLANNING STRATEGIES

- Assessment information is used as basis for planning
- Transition-focused planning begins no later than by age 14
- Meeting time adequate to conduct planning
- Preparation time adequate to conduct planning
- Transition planning meeting time and place conducive to student and family participation
- Process evaluation of planning process relevant to fulfillment of responsibilities
- Multiethnic and multicultural perspective
- Accommodations made for limited English proficiency
- Functional evaluation of student's social abilities
- Functional evaluation of student's cognitive abilities
- Functional evaluation of student's physical abilities

Interagency Collaboration

INDIVIDUAL-LEVEL PLANNING

- Individual transition planning team includes student, parents, school personnel, and appropriate related or adult services personnel
- Student- and family-centered approach to planning and service delivery
- Agency contact with student occurs prior to student's exit from school
- Referral to adult service provider(s) occurs prior to student's exit from school
- Individual transition team leader identified

COLLABORATIVE SERVICE DELIVERY

- Duplicative services reduced
- Duplicative requests for information reduced
- Delineated fiscal resource sharing
- Reduction of system barriers to collaboration
- Collaborative funding of transition services
- Delineated personnel resource sharing
- Collaborative use of assessment data
- Coordinated delivery of transition-related services
- Program information disseminated among cooperating agencies
- Shared delivery of transition-related services or training
- Collaborative planning and service development

ORGANIZATION-LEVEL PLANNING

- Collaborative consultation between special, "regular," and vocational educators
- Transdisciplinary policies and procedures related to transition
- Projection of upcoming service needs
- Transdisciplinary student assessment requirements and processes
- Collaboration between post secondary education institutions and the school district
- Interagency coordinating body includes consumers and family members
- Annual evaluation of interdisciplinary policy and procedures
- Ongoing community-level planning focused on transition-related issues and services
- Community resource directory
- Business and industry involvement in program development

INTERORGANIZATIONAL FRAMEWORK

- Existence of interagency coordinating body
- Formal interagency agreement
- Interagency coordinating body includes employer representation
- Roles of agencies related to transition service delivery clearly articulated
- Established methods of communication among service providers
- Student information shared among agencies (with appropriate release of information and confidentiality)
- Established procedures for release of information among agencies
- Single-case management system
- "Lead" agency identified
- Designated transition contact person for all agencies

HUMAN RESOURCE DEVELOPMENT

- Transdisciplinary staff development activities
- Training activities for employers
- Training activities focused on student and parent empowerment

Family Involvement

FAMILY INVOLVEMENT

- Parent/family participation in evaluation of community-level transition planning
- Parent/family participation in policy development
- Parent/family participation in program evaluation
- Parent/family participation in service delivery
- Parent/family involvement in student assessment
- Parent/family participation in evaluation of individual-level transition planning
- Parents/families exercise decision making
- Parent/family attendance at IEP meeting
- Active parent/family participation in planning process
- Parents/family members as trainers
- Parents/family participation in staff development
- Parents/family members as mentors
- Parents/family role in natural support network
- Parents/family members as volunteer service providers
- Parent/family responsibilities relative to transition planning specified

FAMILY EMPOWERMENT

- Pre-IEP planning activities for parents/families
- Parents/families presented with choices
- Transition information provided to parents/families prior to student's age 14
- Structured method to identify family needs
- Parent/family support network
- Provision of interpreters
- Child care for transition-related planning meetings (e.g., IEP, ITP)
- Respite care
- Flexible planning meeting times
- Flexible meeting locations
- Directory of transition services
- Information to parents/families provided in their ordinary language

FAMILY TRAINING

- Parent/family training re: promoting self-determination
- Parent/family training re: advocacy
- Parent/family training re: natural supports
- Training for parents/families focused on their own empowerment
- Parent/family training re: transition-related planning process (e.g., IEP, ITP)
- Parent/family training re: agencies and services
- Parent/family training re: legal issues

Program Structure and Attributes

PROGRAM PHILOSOPHY

Education provided in least restrictive environment
Integrated settings
Accessibility to all educational options (secondary and post secondary)
Transition outcomes and issues infused in all curricular areas
Outcome-based curriculum
Functional curriculum
Cultural and ethnic sensitivity
Consumer-directed programming
Flexible programming to meet student needs
Outcome-based planning
Longitudinal approach to transition (early childhood to adult)

PROGRAM EVALUATION

Student follow-up
Student follow-along
Data-based management system
Evaluation utilization for program improvement
Ongoing program evaluation
Evaluation of student outcomes
Student/family role in program evaluation
Secondary-level education services needs assessment
Post-school services or program needs assessment

STRATEGIC PLANNING

Regional-level strategic planning
State-level strategic planning
State-level transition body focused on state issues and services
Regional-level transition body focused on regional issues and services
Community-level transition body focused on local issues and services
Community-level strategic planning

PROGRAM POLICY

Adult service systems restructured to include transition-related planning and services as integral components
Coordination between secondary and post secondary education programs
Administrative, school board, and community support for the program
Mission clearly articulated
Values clearly articulated
Shared principles within interagency system
Consistent policies between and within agency and education participants
Transition planning program structure and process clearly articulated
Education system restructured to include transition-related planning and services as integral components

HUMAN RESOURCE DEVELOPMENT

Transition practices resource materials available to personnel
Qualified staff
Preservice training re: transition practices
Ongoing staff development
Disability awareness training
Sufficient allocation of personnel
Technical assistance re: transition practices and planning
Establishment of transition-related personnel competencies

RESOURCE ALLOCATION

Creative use of resources
Multiple utilization of funds
Sufficient allocation of resources
Student/family role in resource allocation
Resources transferred from sheltered and or segregated facilities to community-based and/or integrated settings

15. Please provide a description of how the transition practice(s) has been implemented. The description should be detailed enough to provide an understanding about what is being done, how the service or instruction is delivered, where the service or instruction is provided, to whom the service or instruction is directed (participants), who is providing the instruction or service, and the sequence of events.

Attach additional sheets if necessary.

16. Please indicate how the effectiveness of the practice, program, and/or strategy has been evaluated:

______ External evaluation consultant
______ Case study(ies)
______ Quantitative experimental design
______ Qualitative experimental design
______ Single subject research
______ Other ______________________________

17. Please identify the outcomes for which evaluation data were collected (e.g., employment rates, self-esteem, self-determination skills, etc.)

18. For each targeted outcome, please describe the evaluation findings. The description should be detailed enough to indicate who, what, where, when, and how the evaluation occurred and the findings that were reported.

Please attach additional sheets if necessary.

19. Please attach any relevant products, reports or supplemental materials that would provide information about the transition practices and strategies and the evaluation results.

Please return the nomination form and supplemental materials in the envelope provided by September 30, 1994 to:

Paula D. Kohler, Ph.D.
Transition Research Institute
University of Illinois
113 Children's Research Center
51 Gerty Drive
Champaign, IL 61820

Thank You For Your Participation!!

APPLICATION

1996 EXEMPLARY CAREER GUIDANCE AND COUNSELING PROGRAM SEARCH

SPONSORED BY
U.S. Department of Education Office of Vocational and Adult Education (OVAE)
National Center for Research in Vocational Education (NCRVE)

IN COOPERATION WITH
National Association of State Career Development/Guidance Supervisors (NASCD/GS)
and American Vocational Association (AVA)

PART 1: GENERAL INFORMATION (2 POINTS)

Title of Program: ______________________________

____

Program Contact: ______________________________

Position: ______________________________

Institution/Agency: ______________________________

Program Address: ______________________________

____

City: ______________ State: __________ Zip: __________

Program Phone: (______)____________ Ext.________ Fax: (______)____________

Email Address: ______________________________

I certify that the information included in this application is accurate to the best of my knowledge.

Name of Program Coordinator *(Please print or type.)*	Name of Chief Administrator of School/College *(Please print or type.)*
Signature	Signature
Title	Title
Date	Date

It is important that you have the endorsement of the state official responsible for career guidance programs at your level (middle and secondary programs, community college). Please attach a letter of support from your state official responsible for career guidance programs. (While all state public schools–K-12–have a state person overseeing the career guidance programs in the state, this may not be true of programs in other settings.)

(This page should be completed and returned with the rest of the application. See page 3 for mailing instructions. Completed applications must be postmarked no later than March 15, 1996.)

PART 2: DEMOGRAPHIC INFORMATION (8 POINTS)

This form (Part 2) should follow the cover page of the application.

Title of Program: ____

____

Number of Students/Clients Served During 1994-95 School Year: ____________

Does the total represent ALL students in your school/institute? ___Yes ___No

On the left, please check the different categories of students your program serves and on the right, give an estimate of their numbers.

		Number Estimate
_____	Disadvantaged	_____
_____	Dropouts	_____
_____	Students with disabilities	_____
_____	Minorities *(specify)*	
	____________________	_____
	____________________	_____
	____________________	_____
	____________________	_____
_____	Displaced homemakers	_____
_____	Incarcerated/offenders	_____
_____	Single/teen parents	_____
_____	Dislocated worker	_____
_____	Non-traditional enrollees	_____
_____	LEP/immigrants	_____
_____	Other *(specify)*	_____
	________________________	_____
	________________________	_____

Please indicate the numbers of students who plan to

_____	Plan to attend a 2 or 4 year postsecondary school (if secondary school is applying)	_____
_____	Are enrolled in a vocational/technical program	_____
_____	Other *(specify)*	_____
	________________________	_____
	________________________	_____

(This page should be completed and returned with the rest of the application. See page 3 for mailing instructions. Completed applications must be postmarked no later than March 15, 1996.)

Present Setting:
(choose one)

____ Comprehensive High School
____ Specialized Vocational High School
____ Secondary Level Area Vocational Center
____ Alternative Secondary School (specify): ____________________
____ Postsecondary/Technical Institute/Community College
____ Alternative Postsecondary Program (specify): ____________________
____ Adult and Continuing Education Agency
____ Other: ____________________

(choose one)

____ Public
____ Private

Title of Program Setting *(e.g., Urbana High School):* ____________________

Number of Career Guidance and Counseling Program Staff:

Number

____ Administrator(s)
____ Counselor(s)
____ Job Placement Coordinator(s)
____ Clerical Support Personnel
____ Other: ____________________
____ Other: ____________________

Ratio of Counselors to Students: #________ / #________
Counselors *Students*

1994-95 Budget: $______________ Total Program Budget

Budget Breakdown:

$________ Federal Contribution	$________ Business or Industry Contribution
$________ State Contribution	$________ Other: ______________
$________ Local Contribution	$________ Other: ______________

IMPORTANT

Evidence of Program Effectiveness:

On a separate page, please provide a narrative of the evidence of the effectiveness of your program for **1993-1994 and for 1994-1995.** Focus on the benefits of the program -- what changes occurred because of the career guidance program, what gains were made by the students, how these gains were demonstrated, what difference the program has made. Examples of evidence are career development competencies demonstrated by students (self assessment, career decisionmaking, career planning, etc.), success rates of dropout programs, numbers of students employed upon graduation, completion rates, numbers in postsecondary education, all follow-up data, follow-along information to document program effectiveness, etc. Please append supporting documentation (e.g., Annual Reports, summary of follow-up data).

(This page should be completed and returned with the rest of the application. See page 3 for mailing instructions. Completed applications must be postmarked no later than March 15, 1996.)

PART 3: PROGRAM INFORMATION (4 POINTS)

Program Abstract (2 points)

The abstract should be a ***single page description*** of the program. The abstract should describe the mission and legal status (e.g., school district, nonprofit organization) of the applicant agency, how long the program has been in existence, the purpose of the program, how that purpose relates to the needs of the populations served, the goals of the program, the method used to achieve those goals, and the observable outcomes attained by program participants.

Program Operation (2 points)

Provide a ***single page description*** of how the program operates. The description of the program operation should include program staffing patterns and responsibilities, the number and characteristics of program participants served, services offered, links between program completion and outcomes attained, the scope of the program, instructional approaches and materials used, program innovations, and types of collaborative agreements.

On a separate page, provide a scenario of what a typical day is like at the program setting. Describe what the students and staff are doing at certain periods during the day.

PART 4: COMPONENTS OF EXEMPLARY CAREER GUIDANCE AND COUNSELING PROGRAMS (86 POINTS)

Please describe your program in relation to the following components of exemplary career guidance and counseling programs. These components have been identified through (a) a literature review of career guidance and counseling programs, (b) research on exemplary schools and exemplary instruction, (c) a review by leaders in the field, and (d) provisions of current federal legislation.

A. Career Guidance and Counseling Program Plan

This section describes the heart of the Career Guidance and Counseling Program. It should be comprehensive and include the process, resources, and materials used to assess participants' vocational/career interests and abilities, and how that assessment is utilized in individual program planning. ***Cite, number, title, and attach relevant documents as appendices.***

1. Assisting students/clients to achieve career development competencies (**33 points total**)

Consistent with the age and/or maturity level of your clients/students, describe how the program addresses each of the components below (1.1 - 1.3). You may wish to consult Appendix A for the competencies recommended in the National Career Development Guidelines (NOICC, 1989). While it is not required that those competencies be used in this section, it is important that this section be comprehensive. ***Use no more than two pages for each of the three following areas (A. 1.1 - 1.3).***

1.1. Assist students/clients to increase self-knowledge and self-advocacy (**11 points**)

1.2. Assist students/clients in educational and occupational exploration (**11 points**)

1.3. Assist students/clients in career planning, preparation, and transition (**11 points**) ***(Describe how each student's individual career plan is developed.)***

(The Career Information Delivery System and the Assessment Program should be addressed in 1.1, 1.2, and 1.3.)

All statements made in each of these components (A. 2 - 3) should not exceed 200 words.

2. Addressing the Needs of Diverse Student Populations

   Please describe the participants served by the program, and how the program addresses their special needs. This description should detail the needs and background of program participants, and how that diversity is represented and fostered in the program. Please include how gender differences are addressed. **(6 points)**

3. Program Support Services

   Please describe the support services utilized within the program to meet the purpose and goals of the program. This section can include any resources, special materials, and/or personnel (e.g., aides, volunteers, peer counselors) involved in the program that uniquely assist in achieving its goals. **(6 points)**

B. Collaboration, Articulation, and Communication

All statements made in each of the following components (B. 1 - 4) should not exceed 200 words. Cite, number, title, and attach relevant documents as appendices.

1. Family/Parental Involvement and Support

   Please describe how this program involves the parents and families of participants, if applicable (e.g., secondary schools programs). This description should include information regarding the involvement of parents in (a) general program planning and development, (b) planning for their children, and (c) an advisory role. Also, explain how both students and parents (secondary schools only) are notified of vocational opportunities (required by the 1990 Carl D. Perkins Vocational and Applied Technology Education Act). **(5 points)**

2. Faculty/Staff Involvement in Career Guidance and Counseling Program

   Please describe how academic and vocational educators are involved in the career guidance and counseling program. ***Cite and attach relevant documents and planning forms as appendices.*** **(5 points)**

3. Intra- and Interagency Collaboration

   Please describe both the intra- and interagency collaboration arrangements developed and maintained by program staff. The description of intra-agency cooperative arrangements should include: (a) departments and programs within the educational institution which provide support services, resources, and general assistance to the diverse student/client populations and to the staff of the career guidance and counseling program; (b) the coordination activities conducted; and (c) the benefits of this collaboration. The description of the interagency cooperative arrangements should: (a) name the external agencies and organizations which provide assistance to program staff and/or participants; (b) detail the services and resources provided; and (c) outline the benefits occurring from these collaborative efforts. The guidance program's advisory council should also be addressed. ***Cite and attach any appropriate cooperative agreements.*** **(5 points)**

4. Collaboration with Business

   Describe the collaborative efforts between the program and area businesses or industry. **(5 points)**

C. Institutional Support, Leadership, and Program Evaluation

All statements made in each of these components (C. 1 - 7) should not exceed 200 words. Cite, number, title, and attach relevant documents as appendices.

1. Institutional Support

   Describe how the administration (e.g., president, principal, superintendent) and/or governing body (e.g., school board) support the career guidance and counseling program. Describe any policies that support the program, including policies concerning ratios of counselors to students/clients. **(3 points)**

2. Facilities

   Describe the facilities where the program is housed, including areas for conducting group activities, if available. **(3 points)**

3. Financial Support

   Present a brief explanation of the costs associated with the program. Also, list the funding sources for the program. If the program is receiving special funding (e.g., state grant) ***please identify what revenues will be available to continue the program when the special funding ends.*** **(3 points)**

4. Guidance Personnel Qualifications

   Describe the background of all counselors in the program. Include their work experience, education, and professional credentials or licensing. Include a description of staffing patterns for the program. **(3 points)**

5. Professional Development

   Please describe the professional development activities utilized or conducted by the program. This description should include any preservice, inservice, and/or continuing education obtained by program staff. Additionally, provide a description of the professional development activities conducted by program staff for others such as peer counselors, faculty, or educational agencies. **(3 points)**

6. Program Evaluation

   Describe the evaluation procedures utilized by the program. This description should include information on how evaluation procedures are conducted, who leads the evaluations, those who have access to the evaluation results, and how those results are used to guide the program. ***Cite and attach any available evaluation reports as appendices.*** **(3 points)**

7. Follow-Up of Program Completors and Noncompletors

   Describe the data and information collected by program staff from program graduates/completors and those who do not complete the program. This description should include: (a) the procedure utilized by program staff to collect follow-up information; (b) how that information is analyzed and reported; and (c) how that information is used to improve the program and services. ***Cite and attach previously conducted follow-up reports as appendices. Please do not send raw data.*** **(3 points)**

(See page 3 for mailing instructions. Completed applications must be postmarked no later than March 15, 1996.)

PEPNet Self-Nomination Form

This form must accompany your application. Please place it on top of your submission.

Name of Initiative/Program/Activity ____________________

Name of legal entity responsible for Initiative/Program/Activity ____________________

____

Name of person submitting application ____________________

Title ____________________

Address ____________________

Telephone number ____________________

Fax number ____________________

E-mail address ____________________

Number of pages attached ____________________

(Maximum of 25 pages, including attachments [exclusive of evaluations -- per section 5e of application])

I hereby nominate the above initiative to become part of **PEPNet.** *I verify that the information provided in this application is true. I also agree that this information is available for public dissemination. If my initiative is selected to be recognized, I agree to make its most recent program and/or fiscal audit available on request, and I agree to participate in* **PEPNet** *workshops and training sessions and make my initiative available for site visits so that others may learn from my initiative.*

Signed: ____________________ *Date:* ____________________

Applications to be sent to:
National Youth Employment Coalition
1001 Connecticut Ave., NW, Suite 719
Washington, DC 20036

PEPNet Application -- Questions

Please answer all the questions and provide the necessary information, using the formatting guidelines outlined in the general instructions above. Total number of points for each section is in parenthesis.

1. Purpose and Activities *(0 points, but related to all other sections and scoring)*

Successful initiatives have clear and well-understood aims, and a coherent and plausible set of components and activities to attain them.

(a) Describe the specific mission/aims of the initiative.

(b) How are these aims communicated to staff and youth in the initiative?

(c) Describe the youth who are served by the initiative, in terms of age, income status, education level, residence, etc. Why have these youth been selected?

(d) What specific activities or services are used in this initiative? Which are viewed as essential to the initiative?

(e) Why are these particular activities/services employed? How do they support or further the aims you are seeking to achieve?

(f) How does the organization's overall mission relate to the youth initiative described here?

(g) Describe the relationship between the program and the community.

2. Organization and Management *(50 points)*

Successful initiatives have engaged leadership, qualified and committed staff, work in collaboration with others, and use information and data to continuously improve.

(a) How long has the organization been in business?

(b) How long has the initiative been in existence?

(c) Describe the continuity of leadership.

(d) List the minimum qualifications for key professional staff.

(e) Describe the agency's commitment to staff development. Include a brief description of staff development activities and how these activities assist in meeting the mission of the initiative.

(f) What are staffing patterns and ratios of the youth initiative?

(g) How is information about youth obtained, kept, and utilized by staff? List the types of information.

(h) What information about operations and outcomes is collected for management purposes? How is it used?

(i) What opportunities are there to assess the initiative and gauge how well it is working? Who participates? How readily does the initiative adapt as needed? How is the staff involved in tailoring the initiative?

(j) What collaborative ties (formal and informal) are developed and maintained with other organizations? Please describe. Who is responsible for linking with other organizations? List the ways in which the linking is accomplished.

(k) List the sources of support including the number of years the support has been given.

3. Youth Development *(50 points)*

Successful initiatives are characterized by conscious and professional reliance on youth development principles to identify or shape program activities and to drive the kinds of outcomes that are sought for young participants.

(a) List the standardized assessments which are conducted and the purpose of each. Please submit a copy of the forms or documents you use.

(b) How many adults work with youth in the initiative? Describe their role. Are there built-in opportunities for adults and young people to interact informally? Please describe.

(c) Do youth have a role in the governance of the initiative or institution? Do they have other chances to exercise responsibility for parts of the initiative, or their participation in it?

(d) Describe program activites that promote cultural/ethnic awareness.

(e) What opportunities are there for young people to make suggestions to improve or modify program activities? How does the program respond?

(f) Describe how the initiative's schedule is organized around young peoples' needs and interests.

(g) How does the initiative tailor its activities to the different ages, maturity levels or readiness of individual youth?

(h) How does the program involve family and peers?

(i) Describe additional services and supports you provide to young people in your program.

(j) How are young people recruited for the initiative?

4. Workforce Development *(50 points)*

Successful initiatives consciously prepare youth for the workforce, connect successfully to employers, make appropriate use of training, workplace exposure, work experience, education, and consistently emphasize the connection between learning and work.

(a) What workforce development activities or service elements are incorporated in your initiative (e.g., job readiness, skills training, career guidance, career awareness, career exploration, community service, work experience, summer jobs, etc.)? Please describe.

(b) Describe the linkages the initiative has with employers. What roles do they serve (e.g., employers, trainers, advisors, etc.)?

(c) How do you document workforce development attainments of youth? Do you use any national listings of competencies and standards as part of this documentation? Which? If not do you use other assessments or standards? Please describe or append.

(d) How does your initiative tailor workforce-related activities to youth of different ages, readiness or maturity levels?

(e) How much and what kind of support do you provide to youth on the job/in further education and training, or after they have finished the formal part of your initiative? Do you track youth after they have left? For how long? And how so?

5. Evidence of Success *(50 points)*

Successful initiatives collect and make use of credible data or other measures that reflect the soundness of their goals, their operational effectiveness, and their ability to achieve desired outcomes.

(a) List your expected outcomes. How do you measure the outcome(s)?

(b) Are there outcomes you believe you are achieving but have not measured (please specify)?

(c) What are the costs associated with the initiative and how many youth are served?

(d) Provide copies of your initiative's results (at minimum for the most recent program year).

(e) Have you ever had an outside evaluation? Please attach a full copy of the evaluation report (this report will not count against the 25-page limit for your application).

Appendix C: Sample Rating Forms

SPECIAL EDUCATION EXEMPLARY PROGRAMS REVIEW
IDAHO PUBLIC SCHOOLS

District Name and Number ____________________ Program _____________ Date _______

DOMAIN	#Actual #Possible	Unique Features Weighting	Total Score
A. STUDENT OUTCOMES *Comments:*	/35	/5	/40
B. LEAST RESTRICTIVE ENVIRONMENT *Comments:*	/30	/5	/35
C. COLLABORATION *Comments:*	/20	/5	/25
D. INSTRUCTION *Comments:*	/20	/5	/25
E. PARENT/COMMUNITY INVOLVEMENT *Comments:*	/20	/5	/25
F. PERSONNEL *Comments:*	/15	/5	/20
G. RESOURCES *Comments:*	/10	/5	/15

TOTAL SCORE __________
(185 possible)

SCORING

District ______________________

	# Actual # Possible	Unique Features Weighting	Score
A. STUDENT OUTCOMES			
1. Full Participation in a Variety of Activities a. Expectations and Monitoring of b. Skills teaching	/35	/5	/40
Unique Features:			
Comments:			

	# Actual # Possible	Unique Features Weighting	Score
B. LEAST RESTRICTIVE ENVIRONMENT			
1. Successful Inclusion a. Assessment b. Emphasizing "With Whom" c. Related Services	/30	/5	/35
Unique Features:			
Comments:			

SCORING

District ______________________________

	# Actual # Possible	Unique Features Weighting	Score
C. COLLABORATION			
1. Smooth Correlation of Services a. In-District b. Out-of-District	/20	/5	/25
Unique Features:			
Comments:			

	# Actual # Possible	Unique Features Weighting	Score
D. INSTRUCTION			
1. Flexible Instructional Practices a. Full access to curriculum & materials b. Planning and Assessing c. Proven practices	/20	/5	/25
Unique Features:			
Comments:			

SCORING

District ______________________________

	# Actual # Possible	Unique Features Weighting	Score
E. PARENT/COMMUNITY INVOLVEMENT			
1. Education Programs a. Assisting Parents b. Assisting community	/20	/5	/25
2. Communication a. Parent and family involvement b. Business and community involvement			
3. Involvement in Program Evaluation a. Opportunities for involvement			
Unique Features:			
Comments:			

SCORING

District ____________________

	# Actual / # Possible	Unique Features Weighting	Score
F. PERSONNEL			
1. Staff has High and Realistic Expectations a. Cooperative work	/15	/5	/20
2. Administration Fosters Improvement a. Effective communication b. Strong leadership role			
3. Personal and Professional Growth a. Staff development			
Unique Features:			
Comments:			

	# Actual / # Possible	Unique Features Weighting	Score
G. RESOURCES			
1. Adequate Resources and Allocation	/10	/5	/15
Unique Features:			
Comments:			

SAMPLE FORM

RATING FORM

Applicant: ______________________________

Signature of Reviewer: ______________________________

Date: ______________

		Total
Part 1: General Information	2 points possible	____
Part 2: Demographic Information	8 points possible	____
Part 3: Program Information	4 points possible	____
Part 4: Components	86 points possible	____
A. Career Guidance and Counseling Program Plan		
1. Assisting students/clients to achieve career development competencies		
1.1. Assist students/clients to increase self-knowledge and self-advocacy	11 points possible	____
1.2. Assist students/clients in educational and occupational exploration	11 points possible	____
1.3. Assist students/clients in career planning, preparation, and transition	11 points possible	____
2. Addressing the Needs of Diverse Student Populations	6 points possible	____
3. Program Support Services	6 points possible	____
B. Collaboration, Articulation, and Communication		
1. Family/Parental Involvement and Support	5 points possible	____
2. Faculty/Staff Involvement in Career Guidance and Counseling Program	5 points possible	____
3. Intra- and Interagency Collaboration	5 points possible	____
4. Collaboration with Business	5 points possible	____
C. Institutional Support, Leadership, and Program Evaluation		
1. Institutional Support	3 points possible	____
2. Facilities	3 points possible	____
3. Financial Support	3 points possible	____
4. Guidance Personnel Qualifications	3 points possible	____
5. Professional Development	3 points possible	____
6. Program Evaluation	3 points possible	____
7. Follow-Up of Program Completors and Noncompletors	3 points possible	____
TOTAL	100 points possible	____

Overall Comments: Please list major strengths and weaknesses/concerns about the program.

____

Overall Recommendation:

____ Do not recommend for site visit/evaluation.

____ Recommend for site visit/evaluation. (Please indicate below your concerns about the program, if any, that need to be addressed during the visit.)

____

Screening Promising Practices For Adoption

YOU ARE THE BUYER. You are the selector, or user of promising practices for classroom use by students, or for the education of teachers, or for making changes in service delivery systems. THE SELLER is the developer and/or purveyor of a promising practice. The seller is responsible for presenting his wares in terms that will make it possible for: (a) knowing what you need programmatically; (b) knowing yourself — your local conditions, contexts, and resources; and (c) asking structural questions whose answers will tell you whether or not a promising practice will fit local characteristics.

WHEN YOU USE THIS SCREENING SHEET, you are shopping to fill an identified programmatic need. (For example, you already know the curricular area and target population to be served by the practice you are seeking.) This Screening Sheet lists structural and contextual questions that you should ask about a practice and about yourself, and it suggests a scale for rating the relative match between the practice and yourself.

	Ask The Seller	**Ask Yourself**	**Rate The Match**
Claims	What goals and objectives is the practice designed to achieve?	What goals and objectives am I seeking to fulfill?	1 2 3 4 5 No Match … Good Match
	Notes ____________		
Effectiveness	What evidence proves this practice is successful in achieving what it claims?	How stringently should effectiveness be demonstrated by practices in this domain? IS THE EVIDENCE CONVINCING?	1 2 3 4 5 No Match … Good Match
	Notes ____________		
Relative Advantage	Does the new practice offer something that makes it better than what is already in operation here?	What evidence is there that shows the need to adopt this or any other new program?	1 2 3 4 5 No Match … Good Match
	Notes ____________		
History of Use	In what administrative, geographical, and educational settings has the practice been used?	What are the definitive qualities of my administrative, geographical, and educational setting?	1 2 3 4 5 No Match … Good Match
	Notes ____________		

(continued)

BEST COPY AVAILABLE

Screening Promising Practices For Adoption

Page Two

	Ask The Seller	Ask Yourself	Rate The Match
Minimum Standards For Replication	Exactly what must the adopter do to achieve success?	What am I willing and able to do to replicate a practice?	1 2 3 4 5 No Match Good Match
	Is the practice completely proscribed? Does it permit adjustments and additions by the adopter?	Am I seeking a flexible or proscribed program to meet this need?	
	Can the practice be tried out on a small scale first?	Is a trial effort desirable as part of the decision-making process?	
	Does anything have to be added at the adopter site?	What am I willing and able to add to the practice I adopt?	
	Notes ______________________ ______________________ ______________________		
Compatibility	How much change will this practice create in parallel systems and programs?	How much overall change do I seek? How far does this practice depart from local established policy and practice?	1 2 3 4 5 No Match Good Match
	What background or level of sophistication does the practice demand of users or participants?	What is the background and level of sophistication of my intended users and participants?	
	What are the specific age levels and learning, behavioral, and other characteristics of the children for whom this practice is designed?	What are the characteristics of the children who are intended to benefit from the adoption of a new practice?	
	Notes ______________________ ______________________ ______________________		
Complexity and Convenience	How complicated is this practice? Is there an elaborate set of procedures with a definite sequence?	What level of complexity can we accommodate in implementing a new practice?	1 2 3 4 5 No Match Good Match
	How much discomfort and reorganization will occur if it is put into practice?	What level of discomfort and reorganization can be tolerated?	
	Notes ______________________ ______________________ ______________________		

(continued)

Screening Promising Practices For Adoption

Page Three

	Ask The Seller	Ask Yourself	Rate The Match
Assistance With Implementation	How is this practice conveyed to new users?	What kinds of training or assistance will I need in order to implement a new practice?	1 2 3 4 5 No Match — Good Match
	What follow-up and problem-solving assistance is given?	Will I need external follow-up and problem-solving from the developer?	
	Notes ______________________________ ____________________________________		
Costs	What is the purchase price of adoption and training?	What budget has been estimated for initiating a new practice?	1 2 3 4 5 No Match — Good Match
	What is the cost of implementing and maintaining the practice?	What budget has been estimated for implementing and maintaining a new practice?	
	What additional resources does the practice require in money, personnel, facilities, equipment, and materials?	What additional resources are available for initiating and implementing a new practice?	
	What learning materials or other materials are required? Where do they come from? How much do they cost? Are they reusable or reproducible?	What budget has been estimated for purchasing, supplementing, and reproducing materials?	
	Notes ______________________________ ____________________________________		
Payoff	How rapidly does the practice achieve its goals?	What is my timeline for producing change?	1 2 3 4 5 No Match — Good Match
	How many individuals can participate simultaneously?	What is the total number of individuals intended to participate in this practice?	
	Does the package include evaluation procedures for measuring success?	What are my criteria for judging the success of my replication of this practice?	
	Notes on cost-effectiveness ______________________ ____________________________________		

The ______________________ practice was screened on ______________________ 19____

by ____

It was judged ☐ worthy ☐ unworthy of further consideration, demonstration, and trial use because

____

____

____

BEST COPY AVAILABLE

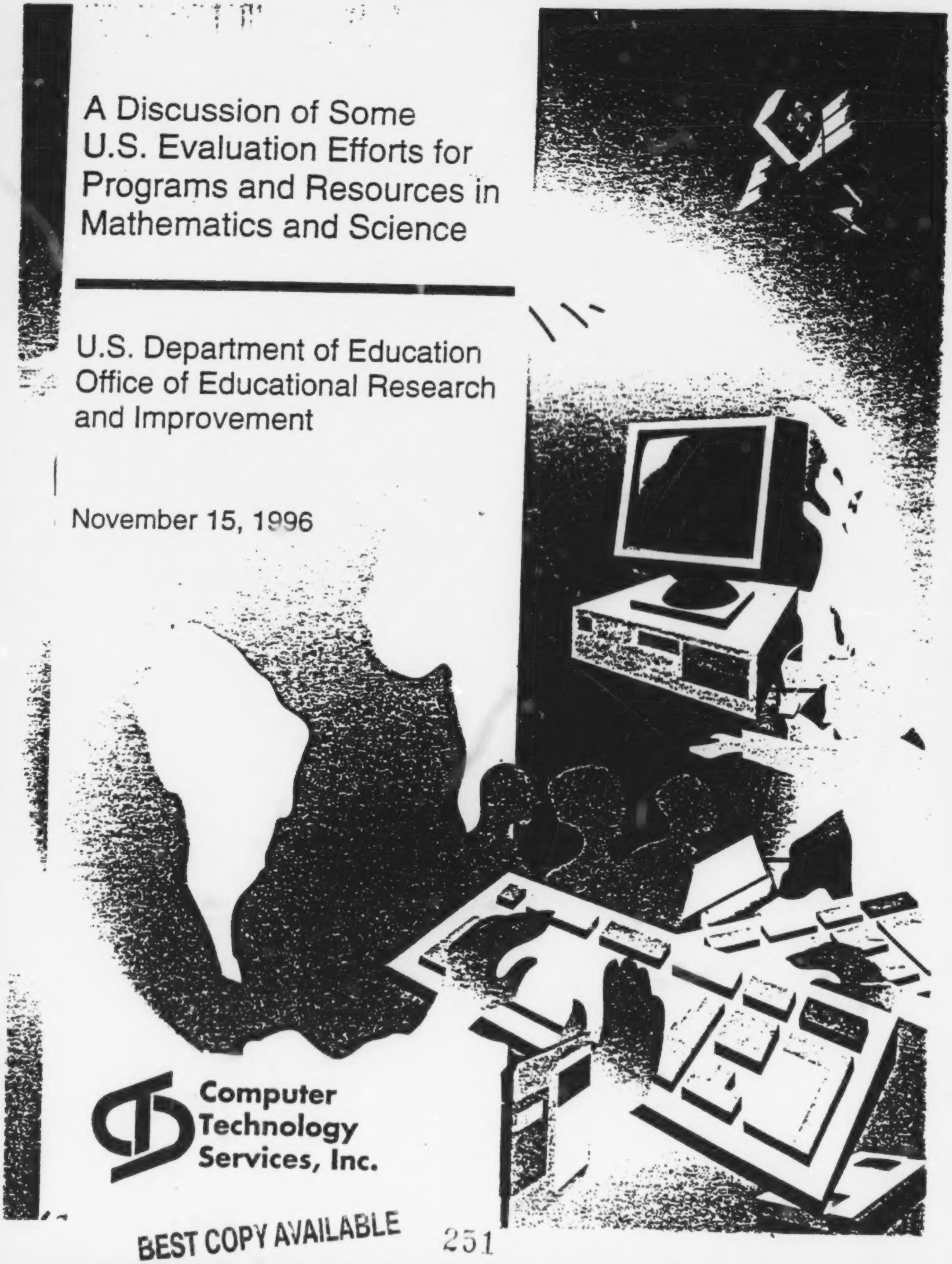
K

250

A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science

U.S. Department of Education
Office of Educational Research
and Improvement

November 15, 1996



Computer
Technology
Services, Inc.

BEST COPY AVAILABLE

A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science

Prepared for:

U.S. Department of Education
Office of Educational Research and Improvement

By:

Carol Muscara
Computer Technology Services, Inc.

November 15, 1996

Table of Contents

Table of Contents

Abstract

A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science

Background and Organization
Collecting and Presenting Information
Identifying Review Efforts
Discussions with Evaluators and Reviewers
"Promising" and "Exemplary"
Difficulties Encountered During Evaluations
Connection with the Math and Science Expert Panel

Bibliography

Appendix A List of Organizations

Appendix B Interview Questions for Evaluators of Programs and Resources

Appendix C Organization Answers to Questions About Evaluation Efforts and Criteria

(1) Matrix for Program Review Efforts
(2) Interview Questions and Organization Responses - Programs
(3) Matrix for Resources Review Efforts
(4) Interview Questions and Organization Responses - Resources

Appendix D Summaries of Information About Evaluating Organizations

Appendix E Sample Evaluation Criteria from Organizations

(1) Programs
(2) Resources

Abstract

A DISCUSSION OF SOME U.S. EVALUATION EFFORTS FOR PROGRAMS AND RESOURCES IN MATHEMATICS AND SCIENCE

by: Carol Muscara
Computer Technology Services, Inc.
1700 Rockville Pike, Rockville, MD 20852
301.468.1160 e-mail: cmuscara@eats.com

ABSTRACT

Fifteen organizations were contacted and asked to discuss their evaluation process for educational products in mathematics and science. Information was gathered from twelve organizations (five for programs and seven for resources) about their evaluation efforts for either programs or resources.

No documented attempt to evaluate policies was found, and organizations did not evaluate practices and programs differently. In analyzing the efforts, all contained the same five evaluation components:

(1) a purpose for the evaluation,
(2) an identified audience for resulting evaluation information,
(3) criteria used to evaluate materials,
(4) a process used to evaluate information, and
(5) evaluation results.

Differences in evaluation efforts depended on the detail needed. Formal, detailed evaluations provided more specific, documented results, while informal, limited evaluations resulted in more general classifications.

All program and resource evaluation efforts looked for quality products with accurate, current content. Most were concerned with correlation to educational research findings and product attention to equity and user appeal. Criteria used by those evaluating programs and resources were similar.

Difficulties encountered by evaluating organizations were the result of funding limitations, time restrictions, inadequate product information, or political considerations.

All organizations offered to share their experiences and results with the U.S. Department of Education Expert Mathematics and Science Panel.

Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science

A DISCUSSION OF SOME U.S. EVALUATION EFFORTS FOR PROGRAMS AND RESOURCES IN MATHEMATICS AND SCIENCE

by: Carol Muscara
Computer Technology Services, Inc.
1700 Rockville Pike, Rockville, MD 20852
301.468.1160 e-mail: cmuscara@eats.com

Background and Organization

In 1994, federal legislation reauthorizing the Office of Educational Research and Improvement (OERI) in the U.S. Department of Education (ED) required that three sets of standards be developed. The standards are intended to guide educators in selecting high quality research and development (R&D) based models or resources developed and tested by others, including those produced with and without federal funding. Legislation stipulates that the standards are to be used by expert panels. Expert panel decisions about what works best then will be shared using national clearinghouses, regional educational laboratories, technical assistance consortia, professional associations, and others.

The second set of the three standards, which is to designate promising and exemplary products (programs, practices, policies, etc.), prompted this paper. OERI commissioned this paper to sample review efforts in mathematics and science education in order to inform the mathematics and science expert panel. The paper contains information and discussion about the process and criteria being used nationwide to evaluate mathematics and science resources, practices, and programs, and how the evaluations are disseminated.

Appendices with information used for this discussion follow the paper. The twelve organizations whose evaluation efforts are summarized are listed in Appendix A. One organization, Project 2061, AAAS, has two resource evaluation efforts summarized, one conducted from 1991-1994 and the second currently in progress. The interview questions in Appendix B were posed to each organization and formed the basis for initial conversation and information gathering. Data from organizations are summarized in three ways, as a matrix and as responses to Appendix B questions in Appendix C, and as a profile of each evaluation effort in Appendix D. Evaluation efforts for programs, and resources are separated, with program efforts in the front of each section. Following the organizational information, Appendix E offers evaluation criteria for two program and two resource evaluation efforts.

Collecting and Presenting Information

After a review of literature with an emphasis on 1990 to the present, a list of twenty potential organizations involved in evaluation and/or review of preK-12 mathematics, science, and technology products was developed. Products (as defined by OERI) include resources, practices, programs, and policies. Five groups were deleted from the original list as the list of organizations was tailored to include as wide a variety of reviewers as possible. The reviewers surveyed include fifteen organizations, foundations, non-profit groups, professional societies, states, regional laboratories, and others. Some of the organizations are not involved with evaluation of programs or resources and are therefore not included in this report. Information from twelve organizations that provide differing approaches to evaluation are reported in this paper.

Questions about the review/evaluation process were compiled before contacting the reviewing organizations. The questions were reviewed and revised with extensive assistance from Dr. Mary Jo McGee-Brown, University of Georgia, Athens. Informal discussions indicated that no organizations evaluated policies and that programs and practices are often considered synonymous. Since programs and resources are dealt with most often by organizations, only those two products were investigated. Programs, for the purpose of this paper, generally include programs and practices. Resources are defined as print and non-print materials useful for learning and/or instruction. A copy of the questions posed to each organization is attached (Appendix B). Questions posed to the organizations also form the vertical axis of each matrix that reports answers from the organizations (Appendix C).

When each organization was contacted, the contact person was given information about the OERI expert panel, the purpose of the interview, and how the collected information would be used. Each contact was asked to discuss the organization's review/evaluation process and, following this input, several additional queries completed the list of questions for each organization. Answers were recorded and summarized for each organization with care being taken to preserve verbatim answers as often as possible. Difficulties encountered during organizational review efforts were also discussed. The same individual conducted all interviews and follow-up.

Initially, two matrices were compiled, one with resource-evaluation strategies, and the second with program-evaluation strategies. These can be found in Appendix C. Most organizations evaluate or review either resources or programs, but not both. An exception was NSTA (National Science Teachers' Association), which reviews both. Each organization type and its purpose for reviewing/evaluating products are in the first two matrix rows. Then, responses to each question are recorded across the page so that viewers can read responses from all organizations. If an organization does not deal with a particular issue, that cell remains blank. Each organization's more complete responses are also included with the questions following the matrices.

The second section of each matrix describes specific criteria organizations use to evaluate programs or resources. Two levels of importance for any criterion are noted. The first shows a high level of interest, indicating that the criterion is essential for an exemplary rating. The second indicates that an organization considers that criterion, but it is not essential for a program or resource to receive an exemplary rating. A check mark in a cell indicates that the criterion is used for evaluation. Since organizations often evaluate or review for special purposes, additional important criteria are sometimes considered. Those criteria are noted below the checklist.

Selections of the detailed criteria used by various groups are included in Appendix E. Two sets of criteria are presented for both program and resource review efforts. While not all organizations publish their selection criteria, those that have, have developed detailed selection items, usually with a specific focus. Broad categories such as content or addressing standards are evaluated using several detailed criteria. For example, for one resource review effort, six criteria are used to reflect the effectiveness of companion teacher materials and seven criteria reflect the quality of science content.

Note: Summaries of information gathered for each evaluation effort were written and shared with the organization to assure content accuracy. The input from organizations is incorporated into the summaries provided in Appendix D.

Identifying Review Efforts

Two methods were used to search for current review efforts. A search of ERIC and other resources at the ED library provided information and references to a significant number of papers about evaluation work, but described few actual evaluation projects. While the library provided informative background material, more was needed.

Searching the World Wide Web (WWW) and other Internet sources for evaluation efforts proved more effective. Many WWW sites provided links to or access information for similar efforts elsewhere. At other sites, electronic conversations often led to additional evaluation efforts. The WWW also provided an opportunity to explore results of many evaluation efforts, giving a flavor of how the evaluations are shared with the education community. The WWW is limited, however, to those organizations with Web presence and access.

As with any identification process, conversations with individuals involved in evaluation efforts led to information about other review efforts that could provide additional insights. If the referenced review effort offered a new perspective, different criteria or process, or an unusual format, it was added to the list to be contacted and included in this paper.

When selecting the organizations to include, we attempted to represent as many kinds of reviewing organizations as possible. We included professional societies, regional laboratories and/or consortia, non-profit and for-profit organizations, and state and local efforts. Two efforts include more than one set of data. NSTA reports on their program and resource evaluation efforts and has a summary in each section. Project 2061, AAAS includes a summary for two resource evaluation efforts: 1) an effort conducted between 1991 and 1994 and 2) an effort in progress. Because ED expressed an interest in technology as it applies to mathematics and science, one review effort that deals exclusively with instructional technology resource evaluation (ISTE) was also included. While the organizations contacted are by no means exhaustive, they represent a variety of groups and a variety of levels of review from informal to formal, from limited to highly detailed. The data for each organization are provided as a written summary (Appendix D). Data are also provided in tabular and question and answer format (Appendix C) that present opportunities for comparison across review efforts.

Discussions with Evaluators and Reviewers

With initial telephone calls to each organization, we attempted to identify the individual most closely connected with and knowledgeable about the evaluation effort conducted by the group. In some cases, we spoke with more than one person about the evaluations. For example, NSTA evaluates programs and resources and reviews resources for different purposes. In all other cases, the organization evaluated either programs or resources, but not both.

Discussions with evaluators indicated that they rarely differentiate among practices and programs. While they agree that there is an actual difference in definition - practices involve something that is done and programs are the results of a group of practices - there is no clear operational delineation between practices and programs. In these efforts, practices and programs seem to run together and the terms are sometimes used interchangeably. For the evaluation process, most evaluators felt that they considered programs and that good practices built good programs.

A thorough data analysis indicated that five components are common to all program and resource evaluation efforts:

- a focus or purpose of the evaluations,
- an identified audience for the evaluation effort,
- criteria used to evaluate,
- the process employed during each evaluation, and
- evaluation results.

Evaluation efforts for resources or programs were quite similar, with differences more noticeable between informal and formal or limited and detailed evaluations. Therefore, the main components of evaluation efforts are compared in the sections that follow by investigating levels of effort for each of the five components rather than whether programs and resources were evaluated.

Focus or Purpose of the Evaluation

As individuals described their evaluation efforts, the focus of each effort seemed pivotal. All other evaluation components, criteria, process, and dissemination depended on the focus for the efforts. For example, if the purpose of evaluations is to develop an educational product, only programs that fit the needs of that product are considered. Or, if a funding agency indicates an interest in a specific area, that area becomes the focus of evaluative consideration. For all efforts that we contacted, the effort's purpose, stated or unstated, drove the evaluation effort.

Informal or limited evaluations: The purpose for informal evaluations might be unstated, but was implied in the work as it progressed.

Formal or detailed evaluations: When organizations engage in formal, detailed evaluations, the purpose or focus is frequently stated (written) as part of the evaluation document.

***Resulting recommendation*: Any organization undertaking evaluation work would benefit by carefully defining its focus or purpose to avoid unnecessary work and too broad a scope.**

Audience for the evaluation effort

By assuming that all the evaluation efforts described here have a mathematics, science, and education orientation, the audience is already partially defined. But subsets of this category include teachers, administrators, students, parents, or product developers. For each evaluation effort, the audience that would use the resulting reviews helped define components for selection. For example, practitioners (teachers, principals) would want to know how resources were implemented in instructional settings, while administrators would be interested in curriculum and standards connections. Parents would find materials with explicit "how-to" instructions valuable.

Informal or limited evaluation: For informal evaluation efforts, a specific audience is rarely identified, and sometimes not even considered. If evaluators are asked to describe the audience, the description tends to be much broader in scope than the evaluation effort would warrant.

Formal or detailed evaluation: While the audience may not be documented, consideration has been given to those who will use the resulting evaluation. The intended audience may be defined as part of the goals statement.

***Resulting recommendation:* Defining the audience for an evaluation effort also helps define the population from which to draw the evaluation sample. Sample parameters inform eventual users about resources or programs they will encounter in an evaluation report.**

Criteria used to evaluate programs and resources

For any evaluation to occur, criteria must be applied to a system. Whether the criteria are as broad as "Is it good?" or as specific as "Is this program/resource free of racial, ethnic, or gender bias?," questions about the content must be asked. Even when there are no stated criteria and only one evaluator, some comparison is taking place during an evaluation. Documented criteria allow more than one person to evaluate, with greater certainty of similar results. The more detailed the documentation, the more reproducible the evaluation. Criteria must be carefully worded to avoid uncertainty and confusion. The most useful criteria for evaluation are offered with examples and discussion among evaluators about the specific wording and meaning of each criterion. Several organizations found that documentation for the parameters of a criterion, as well as training about its meaning, provided the most effective method for producing consistent evaluation. Even with both training and discussion, all personal bias is difficult to overcome. Training for and discussion among evaluators are effective, but quite expensive, undertakings.

Criteria for each evaluation effort are usually the result of combining new development with experience from other evaluation efforts. Teams of experts in all relevant fields for a given evaluation effort often take part in developing the criteria. For example, grade-level experts, content specialists, standards authorities, curriculum writers, publishers, and administrators developed the criteria for one resource effort.

Informal or limited evaluations: These evaluations tend to be quite inclusive of broad resource or program populations, eliminating very few resources or programs. They often have very limited numbers of evaluators. Even if only one person evaluates a series of programs/resources, comparison automatically becomes part of the review process. The intuitive decision that a program/resource is "good" is often the result of this informal selection process.

Formal or detailed evaluation: Criteria for both programs and resources are frequently broken into several categories such as descriptive (title, location of operation, grade level, materials needed) and evaluative (standards correlation, content accuracy, effective use of pedagogical strategies) information.

Descriptive and evaluative categories are often further subdivided into areas such as pedagogical approach, content, design and format, and teacher support. Under each category there are usually four to six evaluation criteria to be applied to a resource or program. Evidence with reference to the program/resource is required for a criterion to be satisfied. Several examples of evaluation criteria for programs and resources are in Appendix E.

In some cases a numeric rating (1 = not addressed, 5 = fulfilled) is applied for each criterion. The resulting total value of all criteria is used to represent how thoroughly a program/resource approaches the highest measure of effectiveness.

Another variable surfaces, especially in more formal evaluation efforts when determining and collecting evidence for how well a program/resource addresses a standard. Because one-time experience with a standard rarely results in learner understanding, the number of times that standard is addressed in a student's history will probably vary. Additionally, the number of times a standard needs to be addressed will vary from student to student. This variability and, therefore, how well a resource or program addresses one or more standards is difficult to capture with criteria.

Criteria used within several program evaluations or resource evaluations are parallel. While the wording of criteria vary among evaluating organizations, the basic content is often similar. A few additions in individual sets of criteria based on project focus or intended audience are evident, but generally the evaluation criteria from effort to effort are parallel. The checklist matrices in Appendix C indicate those criteria that are similar for either resource or program evaluations. In some cases, the criteria are written, in the case of more inclusive evaluation efforts, they might be stated during or implied from the interview.

There are also parallels in criteria for programs with criteria for resources. The similar content for resource or program evaluation criteria is listed below. Those criteria common to all resource efforts and most program evaluation efforts are attention to:

- quality and
- accuracy/currency of content

Common to all but one evaluation effort are attention to:

- a correlation to research findings,
- equity/lack of bias, and
- user appeal

Parallels that exist among criteria for evaluations of programs and resources are:

	PROGRAMS	**RESOURCES**
Descriptive Criteria	program title	resource title
	responsible person with contact information	publisher name with address and contact information
	appropriate age level	appropriate grade level
	instructional time needed	instructional time needed
	average preparation time needed	preparation time needed
	implementation cost/date recorded	cost on a particular date
	audience to be addressed	recommended user
	program description	resource description
	resources needed for implementation	materials needed for implementation

	standards addressed	standards addressed
	articles, references, evaluations	published evaluations, articles
	attention to cost	cost documented with date

Evaluative Criteria	PROGRAMS	RESOURCES
	quality of program	quality of program
	accuracy/currency of content	accuracy/currency of content
	pedagogical effectiveness	pedagogical effectiveness
	correlation with state/national standards	correlation with national content standards
	attention to equity, lack of bias	developed with equity, lack of bias
	multiple content connections	multiple content connections
	developmentally appropriate	age appropriate

Several evaluative criteria were more important to groups reviewing *programs* than those reviewing *resources*. Criteria such as innovation or uniqueness, and transferability to more than one situation were more relevant for program reviewers. On the other hand, ease-of -use, technical/design quality, and built-in user support were important to resource reviewers, but rarely considered when reviewing programs. The importance of community/society involvement is noted in several more recent program and resource evaluation efforts.

The importance of identifying evidence for criteria in the more formal evaluation processes is apparent. Evaluation forms include space for describing the evidence in the program or resource and referencing its location within the materials. The request for evidence is most apparent with the criteria defining a correlation to standards or state frameworks. Evidence is also required whenever change in student or other target population behavior is an evaluation criteria for program success. Both formative and summative evaluations require evidence of behavior change for programs to be considered exemplary.

Resulting Recommendation: **Criteria should be developed by a variety of experts. They should be written in clear language and be described so that users understand each criterion's meaning and purpose. To be most effective, each criterion should be matched to evidence from the resource or program. Using tested, existing criteria for similar evaluations is a logical place to begin when developing criteria for new efforts.**

Process used in evaluation efforts

The process that evaluators follow depends on the formality and scope of the evaluation. The more formal and broader-scope evaluations are developed using many steps for each component of the evaluation process. In all cases however, the process first involves defining a purpose or focus for the evaluation and the audience for whom the evaluation results are intended. As a result, the population of programs or resources that will be evaluated can be defined. Formal or informal criteria then are developed for that evaluation. These four steps seem to be common to all the surveyed efforts, whether formally documented or informally considered. There are instances where one or two of these steps have been overlooked and have been defined at a later stage in the evaluation process, but in all cases, the focus, audience, and sample population to be evaluated can be described and help define criteria for all evaluation efforts.

The methods used to identify and evaluate material vary depending on:

- scope of the evaluation effort,
- funding available,
- time frame for evaluation, and
- formality of the effort.

The broader, well funded, long-term, more formal efforts cast a wide net for material and made use of experts from all associated fields, often using a multi-tiered procedure to produce final choices of exemplary programs or resources. Less structured, informal operations with little funding for evaluation sometimes assign one competent person to perform the entire evaluation process.

Informal or limited evaluations: A few staff members identify programs or resources based on their focus and audience, build criteria or use published existing evaluations, evaluate the material based on printed information, and disseminate the results of their selections. Documentation and evidence for each step are not necessarily collected or archived.

Formal or detailed evaluations: Based on a predetermined focus and audience, criteria are developed by a team of varied experts, tested, revised, and often published. Materials are solicited from a wide variety of sources, and they often receive preliminary screening at the collection level before being submitted for central, more thorough evaluation. For example, states and regional labs use their own criteria to select the best programs from their areas to be submitted for national recognition. An expert group then selects the best from all the submissions for national honor.

Teams of evaluators, most often with varied practitioner expertise, are identified. For example, when K-12 practitioners, university educators, and content specialists evaluate programs, a more comprehensive perspective results. They usually spend time learning about the criteria and their meaning, and then reach consensus about how the criteria will be applied. Evaluations of programs or resources are done initially on the basis of print information in several ways: (1) several individuals evaluate independently and then reach consensus about the evaluations or (2) a group eliminates materials using a multi-tiered process, resulting in a few exemplary programs or resources being chosen. Either during the evaluation or after the exemplary ones are chosen, programs and resources are validated in the field with observation. Results are published for or disseminated to the audience of choice.

Resulting recommendation: **Because evaluation efforts are restricted by funding, time, resources, and other considerations, each effort will be different. The more varied the relevant expertise involved, the more complete the evaluation. As long as its focus, audience, and scope are documented, any evaluation can be valuable to others.**

Evaluation Results

Evaluation results are of little value until they are disseminated in some way to an interested audience. Most efforts make their focus and criteria available as part of a publication. Either hard copy or electronically shared databases are becoming a universal dissemination tool. For example, AAAS and Annenberg are making databases available on CD-ROMs. Online dissemination has made this type of information even more accessible. ENC and EPIE provide access to current resource evaluations on WWW sites. A teacher-evaluation effort conducted on science resources by Ohio will be available on the ENC WWW site November 1996. Local evaluation efforts are receiving more and wider attention with development of school-system World Wide Web pages, but many efforts still remain unavailable.

Evaluations of programs and resources are time dependent. While there is an assumption that exemplary materials will age well, to be effective, their evaluations need to be revisited regularly. As the volume of evaluations increases, it becomes even more difficult (and costly) to update evaluation efforts. To be relevant, evaluation efforts must be published/disseminated with the date they were done. Users can then determine the evaluation's relevance for current needs.

Informal or limited evaluations: Because these are often limited in scope and conducted for a single purpose, the results are most frequently used internally by the evaluating organization.

Formal or detailed evaluations: These evaluations are more heavily funded, often with dissemination as one of several stated goals. Printed versions of the results, with varying quantities of information, are common. Online publication is becoming a major method of communicating results. While many groups publish printed and online versions, some organizations now are publishing only online with no printed copy. Online databases with minimal restrictions should make local evaluation efforts even more accessible.

Resulting recommendation: **Evaluation results need to be disseminated to be valuable. Because they are time dependent, a most-recent-evaluation date is critical for users. The Internet offers rapid, easy access to evaluations in a timely manner. Until online access is universal in the education community, however, printed publication will still be needed.**

"Promising" and "Exemplary"

As these evaluation efforts were discussed with evaluating organizations, there seemed little differentiation between promising and exemplary. Generally, programs with little documentation or resources without instructional trials were considered promising. They look interesting, address learning needs, provide opportunities for good learning, but have no evidence of being effective. Those same promising programs could be considered exemplary after they produce evidence of effectiveness. In other words, they might be exemplary programs or resources, but simply provide no evidence to that effect yet. Promising materials encountered here are actually on a continuum with exemplary materials. Some may be shown to be exemplary, some may be less than exemplary, but all exemplary programs or resources were at one time promising.

Transferability is desirable, but not always essential for the designation of exemplary. Discussions indicate that because all instructional environments, teachers, and student groups vary in content and context, no program or resource will be transferred without modification. There then follows the question of whether the modified, transferred program or resource is the same under those circumstances.

Difficulties Encountered During Evaluations

Funding, time, adequate resources, and political considerations seem to be the major problems encountered by these evaluation efforts, especially the more extensive efforts. In the instance of small, in-house evaluations for a limited purpose, the efforts usually proved effective and encountered few difficulties. In the case of more comprehensive efforts, these four limitations affected results:

(1) Funding was needed to purchase materials, convene and train experts, spend more time doing the evaluation, more thoroughly validate results with practitioners, and disseminate results freely and easily.
(2) Time was needed to better identify the full population of materials, more thoroughly involve practitioners who already work full time, and test materials in various instructional settings.
(3) Adequate resources, including complete material information, computer and online facilities, simultaneous access to all components needed for successful implementation, and contact persons for discussion are essential.
(4) Evaluations of professional materials that could be detrimental to publishers, professional colleagues, or organizations can limit access to information and

resources in the future. Considering all implications of evaluating materials becomes part of each organizational evaluation program.

Connection with the Mathematics and Science Expert Panel

In all cases, the evaluating organizations listed in this paper will gladly share any process, criteria, and results with the expert mathematics and science panel. There were no specific suggestions about how sharing might occur. Some results are currently published, others are not. All materials collected for this paper have been submitted under separate cover to the U.S. Department of Education, OERI.

An additional evaluation effort that would be valuable to the panel occurred during the summer 1996 at Ohio State University (Eisenhower National Clearinghouse, ENC). All K-4 resources housed in ENC were evaluated by a group of 14-16 teachers. Those results should fit well with the panel interest. Results were not yet available as this paper was being written, but the project coordinator, Mr. Todd Fennimore, Project Coordinator for the ENC/SchoolNet Software Review Project, has indicated his willingness to share that work with the panel. The results of the evaluation effort will be available through the Eisenhower National Clearinghouse WWW site (http://www.enc.org) in late October 1996. Information on the evaluation protocol and process is submitted to ED with this paper.

Bibliography

A DISCUSSION OF SOME U.S. EVALUATION EFFORTS FOR PROGRAMS AND RESOURCES IN MATHEMATICS AND SCIENCE

BIBLIOGRAPHY

American Association for the Advancement of Science, *Benchmarks for Science Literacy,* New York, Oxford University Press, 1993

Alkin, M.C., *Debates on Evaluation,* Beverly Hills, SAGE Publishing, 1990

Alkin, M.C. & E. R. House, "Evaluation of Programs" in *Encyclopedia of Educational Research*, 6th ed., New York, MacMillan Publishing Co., 1992

California State Department of Education, *Guidelines for Science Instructional Materials,* Sacramento, 1992

Center on Evaluation, Development and Research, "Effective Classroom Instruction," *Phi Delta Kappan*, v69 n3 1987

Eisenhower National Clearinghouse for Mathematics and Science, *Collection Development Policy,* Columbus, Eisenhower National Clearinghouse, 1994

Eisenhower Regional Consortia, *Promising Practices in Mathematics and Science Education - 1995*, sponsored by U.S. Department of Education, Washington DC, Government Printing Office, 1996

Federal Register, Volume 61, Number 107, 34CFR Part 701, Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI)- Designation of Exemplary and Promising Programs; Proposed Rule

Frazier, G.G., & R. N. Sickles, *Directory of Innovations in High Schools,* Princeton NJ, Eyes on Education, 1993

Guba, Egon G., "What Have We Learned about Naturalistic Evaluation?," *Evaluation Practice*, v8 n1 P23-43, 1987

Hegleson, S., et. al, *Promising and Exemplary Programs and Materials in Elementary and Secondary Schools – Science,* Columbus, ERIC Clearinghouse, 1990

House, Ernest R., "Trends in Evaluation," *Educational Researcher*, v19 n3 p24-28, 1990

Johnson, J. M., "Software Reviews," *Learning and Leading with Technology,* v23 n1 p56 September 1995

Joint Committee on Standards for Educational Evaluation , *Standards for Evaluations of Educational Programs, Projects, and Materials,* New York, McGraw-Hill, 1981

McAlpine, L. & C. Weston, "The Attributes of Instructional Materials," *Performance Improvement Quarterly,* v7 n1 p19-30, 1994.

National Council for Research in Vocational Education, *What Makes Programs Exemplary?*, Published on WWW site: http://vocserve.berkeley.edu/MDS-856/default.html

National Council for Teachers of Mathematics, *NCTM Evaluation Standards, Standard 11*: Indicators for Program Evaluation, NCTM, published on the Eisenhower National Clearinghouse WWW site, http://www.enc.org

National Science Resources Center, *Resources for Teaching Elementary Science*, National Academy Press, Washington, DC, 1996

National Science Teachers Association, *Criteria for Excellence*, Washington, DC, NSTA, 1990

NorthWest Regional Educational Laboratory, *Promising Practices in Small High Schools*, Seattle, NWREL, 1984

Promising Practices in Mathematics and Science Education: A Collection of Promising Educational Programs and Practices from the Laboratory Network Program, Washington DC, 1994

Ralph, J. & M. C. Dwyer , *Making the Case, Evidence of Program Effectiveness in Schools and Classrooms*, Washington, DC, Government Printing Office, 1991

Research in Teaching and Learning, Volume 4, "Mathematics and Natural Sciences," New York, MacMillan Publishing Co., 1990

Saroyan, A. & G.I. Geis, "An Analysis of Guidelines for Expert Reviewers," *Instructional Science: An International Journal,* v17 n2 p101-128, 1988

Scriven, Michael, "Evaluation Ideologies," in G.F. Madaus, M. Scriven & D. L. Stufflebeam (Eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Boston, MA, Kluwer-Nijhoff

Smith, Nick L. & S. Chircop, "The Weiss-Patton Debate: Illumination of the Fundamental Concerns", *Evaluation Practice,* v10 n1 p5-13, 1989

Stalford, C.B., *Issues in Validating Effectiveness of School Improvement Programs*, Paper presented at the Annual Meeting of the American Educational Research Association, SanFrancisco, 1992

Turnbull, Brenda J., "Research Knowledge and School Improvement: Can This Marriage Be Saved?," Presented at American Educational Research Association meeting, San Francisco, 1992

Walberg, H. J. & R. P. Niemiec, "Validating Exemplary Programs: Methods and Criteria," *Educational Evaluation and Policy Analysis*, v15 n4 p429-36, 1993

Appendix A

List of Evaluation Organizations

A DISCUSSION OF SOME U.S. EVALUATION EFFORTS FOR PROGRAMS AND RESOURCES IN MATHEMATICS AND SCIENCE

LIST OF EVALUATING ORGANIZATIONS

PROGRAMS

The Annenberg/CPB Math and Science Project

Concord Consortium

National Science Teachers Association (NSTA)

Northwest Regional Educational Laboratory (NWREL)

SouthEastern Regional Vision for Education (SERVE)

RESOURCES

American Association for the Advancement of Science - Project 2061
    1) Evaluation effort from 1991-1994
    2) Evaluation effort in progress

California State Department of Education

The Educational Products Information Exchange Institute (EPIE)

Eisenhower National Clearinghouse (ENC)

International Society for Technology in Education (ISTE)

Smithsonian Institution/National Academy of Sciences National Science Resources Center (NSRC)

National Science Teachers Association (NSTA)

Appendix B

Interview Questions for Evaluators of Programs and Resources

INTERVIEW QUESTIONS[1] FOR PROGRAM EVALUATION ORGANIZATIONS

The Department of Education wants to know how different organizations/individuals define promising and exemplary practices/programs/policies/resources and how they are currently evaluating them. Expert panels for different subjects will be convened by the Department to provide a "consumer-oriented" selection of promising and exemplary practices, programs and resources. The criteria for both are outlined in a proposed rule published by the Department in the Federal Register , June 3, 1996, Volume 6, #107. The rule is also available on-line (http://inet.ed.gov/legislation/FedRegister/proprule)

THIS INTERVIEW INTENDS TO COLLECT INFORMATION TO DETERMINE:

- In-depth understanding about how the organization deals with program/practices, practices, and programs
- specifically how the organizations identifies and evaluates them
- whether the organization differentiates among promising and exemplary practices, and
- if so how they differentiate.

Interviews will be used to collect data only. No individual names will be associated with the reporting of data. Data will be used for comparison.

CLASSIFYING THE PROGRAM/ PRACTICE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Program/practices

- What type program/practices does your company/organization examine?

- From what population do you select the program/practices for screening.

- What process do you use to identify the population of materials

- Who is involved in the screening/evaluation process?

EVALUATION CRITERIA

- Do you differentiate among promising and exemplary program/practices

- What criteria are used to consider promising or exemplary programs/practices

- What kind of evidence is needed to include a program/practice?

---

[1] These questions were developed with extensive help from Dr. Mary Jo McGee-Brown, University of Georgia, Athens, GA.

Appendix C
Matrix of Organization
Answers to Questions About Evaluation Efforts and Criteria

Resources

	PROGRAMS	
Questions	**Annenberg 1994** **The Guide**	**Concord Consortium**
Purpose of evaluation/choice process	to disseminate information about efforts that are stimulating fundamental, systemic changes in math and science education	for distance-learning teaching, learning, and development applications
Type of organization	foundation	company
What type program/practices does your company examine?	math and science programs, organizations, conferences	programs that can be used for distance learning
From what population do you select the program/practices for screening.	math and science	existing, proven programs
What process do you use to identify math & science programs, organizations, conferences?	submissions, conversations	publications, conversations, conferences
Who is involved in the screening/evaluation process?	staff	individual staff member
Do you differentiate among promising and exemplary program/practices	no	accept only the best that meet criteria
What criteria are used to consider promising or exemplary?--for programs/practices E=exemplary P=promising		materials must be immediately available (not in draft form)
What kind of evidence is needed to include a program/practice?	meet criteria	1. materials must be available 2. not dependent on special materials 3. useable by an average teacher 4. learning in context
Do you endorse a product or practice?	added to The Guide	use it for new distance learning course
How many people are involved screening each item?		1 staff member

[illegible]

Questions	NSTA Criteria for Excellence-	NWREL	SouthEastern Regional Vision for Education - SERVE
Purpose of evaluation/choice process	to provide exemplary models for science teaching	to provide information on good programs to others with similar interests	to choose and disseminate quality math and science program information nationwide
Type of organization	professional society	regional lab	Eisenhower regional consortium for math and science
What type program/practices does your company examine?	science programs	would be of interest to others in 5 state region	math and science programs
From what population do you select the program/practices for screening.	recommended U.S. programs	programs implemented in 5 state region	math and science programs chosen from submissions nationwide
What process do you use to identify math & science programs, organizations, conferences?	State nominates programs to be evaluated by national panel	1. exemplary 2. promising 3. state recognized 4. grass-roots success	1 .consortium nomination 2. regional selection 3. national review 4. site-visit validation
Who is involved in the screening/evaluation process?	states, national panel	NWREL	1,2. local consortia staff, 3. consortia representative panel 4. SERVE staff, experts
Do you differentiate among promising and exemplary program/practices	yes-3-12 programs chosen annually as exemplary	yes, use proposed ED standards	yes E= 50 programs chosen P=most nominated programs in step 1.
What criteria are used to consider promising or exemplary?--for programs/practices E=exemplary P=promising	E=Highly rated in criteria such as those describing curriculum, goals, students, teachers, administration, community relations. P=others nominated	E = meet ED standards	E = meet criteria, selected as exemplary, validated by observation
What kind of evidence is needed to include a program/practice?	nomination from 50 states + DC + Puerto Rico	state approved or expert recommendations, evaluation data	nomination from regional lab
Do you endorse a product or practice?	listed in Criteria for Excellence book	publish in magazine or on the Internet	published in Promising Practices in Math and Science Education annually
How many people are involved screening each item?	state choice +national panel	varies	1, 2. staff 3. 10 representatives 4. at least 2/visit

	PROGRAMS	
Questions	**Annenberg 1994** **The Guide**	**Concord Consortium**
What kind of record is kept of the screening/evaluations?--paper, computerized database, list only	database	paper converted to electronic
How is the evaluation conducted--(site visit, program/practice use in a classroom, stepwise process)	review of program information	interview
How are decisions made (group, numeric average, discussion, combination)?	discussion	discussion
For how long is one review/acceptance valid? Are the accepted items revisited?	updated twice annually (last update was 1994)	

Questions	NSTA Criteria for Excellence-	NWREL	SouthEastern Regional Vision for Education - SERVE
What kind of record is kept of the screening/evaluations?--paper, computerized database, list only	paper	paper	database, annual publication
How is the evaluation conducted--(site visit, program/practice use in a classroom, stepwise process)	from written submissions nationally	informal except for state approval	group decision, site visit
How are decisions made (group, numeric average, discussion, combination)?	group	varies	
For how long is one review/acceptance valid? Are the accepted items revisited?	new choices annually, previous choices remain		

	PROGRAMS			
Questions	The Guide Annenberg 1994		Concord Consortium	
riteria used to evaluate programs	Informal		Informal	
quality	✓		✓	
accuracy/currency			✓	
pedagogical effectiveness	✓		✓	
correlation to research findings	✓		✓	
correlation with standards or state frameworks	✓		✓ Standard	
equity, lack of bias	✓		✓	
user appeal				
user involvement	✓		✓	
ease-of-use			✓ not dependent on special materials	
multiple content connections	✓			
usefulness to others or in another setting			✓	
cost effectiveness	✓		✓	
Built-in user support	✓			
innovation or uniqueness	✓		✓ innovation powerful	
escribe how some criteria might be more heavily eighted than others			1. Usefulness 2. Strong pedagogy & content	
ny other comments				

Questions	Criteria for Excellence-NSTA		NWREL		SouthEastern Regional Vision for Education - SERVE	
riteria used to evaluate programs	Formative and Summative		Informal			
	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]
quality	✓		✓		✓	
accuracy/currency	✓				✓	
pedagogical effectiveness	✓		✓		✓	
correlation to research findings	✓				✓	
correlation with standards or state frameworks					✓	
equity, lack of bias	✓		✓		✓	
user appeal			✓			
user involvement	✓		✓		✓	
ease-of-use					✓	
multiple content connections	✓					
usefulness to others or in another setting			✓		✓	
cost effectiveness						
Built-in user support						
innovation or uniqueness			✓		✓	
escribe how some criteria might be more heavily veighted than others					1. behavior change 2. authentic assessment	
ny other comments	1. Developmentally appropriate 2. Continued education 3. Society-related 4. Goal related 5. Community connection		1. Goal based 2. Produce lasting change		1. Emphasizes evaluation appropriateness for each program	

Appendix C
Interview Questions
and
Organization Responses

Resources

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS

Project 2061, American Association for the Advancement of Science (AAAS)

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS (a) = answer to question

(1) earlier review effort 1991-94

(2) in-progress effort is building a process for educators to use when evaluating resources

Resources

- What type resources does your company/organization examine?
  (a) (1) print and non-print science, math, technology resources of all kinds
  (2)science curriculum material
- From what population do you select the resources for screening.
  (a) (1) all available print and non-print with an emphasis on non-traditional
  (2) print
- What process do you use to identify the population of materials
  (a) catalogs, reviews, teacher recommendation, conversations, publications
- Who is involved in the screening/evaluation process?
  (a) (1) practitioners, Project 2061 staff
  (2) Project 2061 is developing a process; it does not offer reviews of resources
- Do you look at exemplary and/or promising resources?
  (a) (1) a resource is promising until it is shown to be effective with students

EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) (1) yes, promising resources meet all 12 criteria well, but haven't been tested with students; exemplary resources meet all criteria well and work in a learning environment
- What criteria are used to consider promising or exemplary resources?
  (a) (1) promising resources meet the 12 criteria well, exemplary resources meet criteria well and are effective in a learning environment
  (2) whether content matches *Benchmark* content plus matches seven instructional criteria
- What kind of evidence is needed to include a resource?
  (a) (1) Evidence of effectiveness with students and high marks on criteria indicates exemplary Potential for effectiveness with students and high marks on criteria indicates promising
  (2) evidence that a resource could help students achieve science literacy goals
- Do you endorse a product or practice?
  (a) (1) no endorcement, resource meets the needs of 2061 practitioners at this time
- How many people are involved screening each item?

(a) (1) one or more

- What kind of record is kept of the screening/evaluations?–paper, computerized database, list only

    (a) (1) evaluations are entered into a database

EVALUATION PROCESS QUESTIONS

How is the evaluation conducted–(site visit, resource use in a classroom, stepwise process)

(a) (1) a five step process includes 1) screening, 2) evaluating based on criteria, 3) describing , 4) analyzing the resource for evidence of correlation to benchmarks, and then 5) inputting resultant data into the database.

(2) A four step process is recommended: 1) preliminary evaluation, 2) content analysis, 3) instructional analysis, and 4) building a resource profile. (see summary 2)

- How are decisions made (group, numeric average, discussion, combination)?

    (a) (1) numeric based on criteria rating and evidence of criteria correlation in resource

    (2) Resources must serve content and pedagagical intent of *Benchmarks* or National Science Standards

- For how long is one review/acceptance valid? Are the accepted items revisited?

    (a) (1) Each review is dated and results are reviewed by others as time allows. Resource evaluations are planned to be revisited at least every three years.

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS

Educational Products Information Exchange (EPIE) Institute

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Resources

- What type resources does your company/organization examine?
  (a) All available math and science non-print resources.
- From what population do you select the resources for screening.
  (a) Non-print resources available for use
- What process do you use to identify the population of materials
  (a) Periodical articles, published reviews, conversations, conferences
- Who is involved in the screening/evaluation process?
  (a) EPIE staff members who at the time of contact numbered four
- Do you look at exemplary and/or promising resources? (The questions will vary depending on the subject, but maybe they can be similar?)
  (a) No, EPIE tries to include all resources with the idea that educators might be able to use them.

EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) No
- What criteria are used to consider promising or exemplary resources?

- What kind of evidence is needed to include a resource?
  (a) EPIE includes as many resources as possible
- Do you endorse a product or practice?
  (a) No, all identified materials are included in the EPIE database
- How many people are involved screening each item?
  (a) EPIE staff identifies resources
- What kind of record is kept of the screening/evaluations?--paper, computerized database, list only
  (a) Resource information is available on the EPIE database. Access is purchased by schools/districts.

EVALUATION PROCESS QUESTIONS

- How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)
  (a) no EPIE evaluation; resources are accepted with evaluations from elsewhere
- How are decisions made (group, numeric average, discussion, combination)?
  (a) EPIE includes as many resources as possible

For how long is one review/acceptance valid? Are the accepted items revisited

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS
Eisenhower National Clearinghouse

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Resources

- What type resources does your company/organization examine?
  (a) Math and science resources. An initial emphasis was placed on collecting resources from federally funded projects.
- From what population do you select the resources for screening.
  (a) All math and science resources that are submitted to ENC by developers.
- What process do you use to identify the population of materials
  (a) ENC actively searches for resources developed as the result of federal funding. It also accepts resources from other developers.
- Who is involved in the screening/evaluation process?
  (a) ENC staff including content specialists, graduate students, professional staff.

(a) EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) No, but those considered best are highlighted as part of special publications. Recommendations for the publications are also accepted from their math and science advisory boards.
- What criteria are used to consider promising or exemplary resources?
  (a)
- What kind of evidence is needed to include a resource?
  (a) All math and science resources are accepted.
- Do you endorse a product or practice?
  (a) No
- How many people are involved screening each item?
  (a) Two or more
- What kind of record is kept of the screening/evaluations?--paper, computerized database, list only
  A paper form is maintained and the information is included in the ENC database

EVALUATION PROCESS QUESTIONS

- How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)
  (a) No ENC evaluation is conducted. Reviews and evaluations from other sources are included in the database with the resource information.
- How are decisions made (group, numeric average, discussion, combination)?
  (a)

- For how long is one review/acceptance valid? Are the accepted items revisited?
  (a) New reviews are added to resource information when they are published or become available

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS

International Society for Technology in Education (ISTE)

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Resources

- What type resources does your company/organization examine?
  (a) All electronic resources (print materials may be included)
- From what population do you select the resources for screening.
  (a) Non-print resources
- What process do you use to identify the population of materials
  (a) Submissions from producers, conversations, personal contacts.
- Who is involved in the screening/evaluation process?
  (a) ISTE staff and field-based practitioners.

(a) EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) Published reviews differentiate between promising and exemplary in the "conculsion" and "recommendation" section of the review.
- What criteria are used to consider promising or exemplary resources?
  (a) Published reviews use ISTE evaluation criteria as they consider resources (see Appendix E).
- What kind of evidence is needed to include a resource?
  (a) Only reviews that reflect good or excellent materials are published.
- Do you endorse a product or practice?
  (a) No
- How many people are involved screening each item?
  (a) one reviewer
- What kind of record is kept of the screening/evaluations?—paper, computerized database, list only
  (a) A paper record is maintained. Reviews are published in *Learning and Leading with Technology* (previously *The Computing Teacher*)

EVALUATION PROCESS QUESTIONS

- How is the evaluation conducted—(site visit, resource use in a classroom, stepwise process)
  (a) Resource is provided to reviewer. A hands-on review is done.
- How are decisions made (group, numeric average, discussion, combination)?
  (a) The individual reviewer writes a review.
- For how long is one review/acceptance valid? Are the accepted items revisited?
  (a)

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS

California State Department of Education

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Resources

- What type resources does your company/organization examine?
  (a) print and non-print science materials
- From what population do you select the resources for screening.
  (a) Materials submitted by producers for state adoption consideration
- What process do you use to identify the population of materials
  (a) Materials are submitted by producers
- Who is involved in the screening/evaluation process?
- (a) The California State Department of Education trains educators about the state-framework-based criteria and process used for evaluation
  Any educator may evaluate a resource to be purchased with local funds.

EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) Only desirable or exemplary resources are adopted for use in California. To use state funds for purchase, the material must be on the adoption list.
- What criteria are used to consider promising or exemplary resources?
  (a)The resource meets state criteria. Since even exemplary resources can present imperfect (unequal) results, only those with desirable or exemplary evaluations are considered.
- What kind of evidence is needed to include a resource?
  (a) Written evidence of location in the resource that satisfies a criteria.
- How many people are involved screening each item?
  (a) For state adoption, three to five educators reach consensus, at the local level one person can recommend a resource.
- What kind of record is kept of the screening/evaluations?--paper, computerized database, list only
  (a) Records vary, but lists of adopted resources are published in print and electronically

EVALUATION PROCESS QUESTIONS

- How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)
  (a) Resources are sent to more than one evaluator and they must reach consensus on the evaluation before a material is adopted for state use
- How are decisions made (group, numeric average, discussion, combination)?
  (a) Consensus among several trained evaluators must be reached for state adoption

- For how long is one review/acceptance valid? Are the accepted items revisited?
  (a) There are approximately six years between content adoptions at the state level.

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS

NAS/Smithsonian National Science Resources Center

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Resources

- What type resources does your company/organization examine?
  (a) Science materials in print
- From what population do you select the resources for screening.
  (a) All print science resources that support inquiry and hands-on learning. No text books or trade books are considered.
- What process do you use to identify the population of materials
  (a) Letters of invitation are sent to potential producers. Submitted materials are considered for evaluation
- Who is involved in the screening/evaluation process?
  (a) Practicing educators (for pedagogical criteria) and scientists (for content criteria)

EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) Yes, only the most highly rated are included in *Science Resources for Teaching Elementary School Science.*
- What criteria are used to consider promising or exemplary resources?
  (a) Evaluation criteria for pedagogy and content.
- What kind of evidence is needed to include a resource?
  (a) Items/concepts that correlate to the criteria are identified in the resource and noted on the evaluation form.
- Do you endorse a product or practice?
  (a) Resources with exemplary evaluation ratings are included in the publication, *Resources for Teaching Elementary School Science*
- How many people are involved screening each item?
  (a) Two educators (pedagogy), and one scientist (content)
- What kind of record is kept of the screening/evaluations?--paper, computerized database, list only
  (a) Print records of evaluations are retained.

EVALUATION PROCESS QUESTIONS

- How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)
  (a) Educators are convened for several weeks, trained on criteria and asked to review a series of resources. Those resources reviewed highly are sent to scientists for a content review.
- How are decisions made (group, numeric average, discussion, combination)?

(a) Positive reviews in areas of pedagogy and content are required to include reviews in *Resources for Teaching Elementary School Science*.

- For how long is one review/acceptance valid? Are the accepted items revisited?
  The last edition of this resource guide was published in 1988 (*Science for Children)*

INTERVIEW QUESTIONS FOR RESOURCE EVALUATION ORGANIZATIONS

National Science Teachers Association

CLASSIFYING THE RESOURCE SELECTION PROCESS

GENERAL FAMILIARITY QUESTIONS

Resources

- What type resources does your company/organization examine?
  (a) U.S. published science resources (some high school math considered)
- From what population do you select the resources for screening.
  (a) Pre-K-12 science only, no health materials. Only those currently in print are considered. No promotional materials are considered. The materials must emphasize real science with care to avoid pseudo-science and superficial science and must be generally useful.
- What process do you use to identify the population of materials
  (a) NSTA maintains two panels of reviewers, one for print and one for non-print resources. Good reviews are used as indicators of potentially useful materials.
- Who is involved in the screening/evaluation process?
  (a) NSTA staff and 200-300 educators nationwide with expertise in the areas covered by the resource.

EVALUATION CRITERIA

- Do you differentiate among promising and exemplary resources
  (a) Only exemplary resources are sent for review or retained for resale in the NSTA conference stores.
- What criteria are used to consider promising or exemplary resources?
  (a) Exemplary resources meet the criteria in reviewer's judgment.
- What kind of evidence is needed to include a resource?
  (a) Reviewers with specific details of exemplary qualities are used.
- Do you endorse a product or practice?
  (a) Those resources considered useful for teachers and exemplary are purchased for resale in NSTA convention stores
- How many people are involved screening each item?
  (a) one or more and NSTA staff
- What kind of record is kept of the screening/evaluations?--paper, computerized database, list only
  (a) Written reviews are retained.

EVALUATION PROCESS QUESTIONS

- How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)
  (a) For published reviews, materials are distributed to reviewers in the field.
- How are decisions made (group, numeric average, discussion, combination)?

(a) For materials that are resold, several good reviews and a secondary review by NSTA staff are necessary.

- For how long is one review/acceptance valid? Are the accepted items revisited?
  (a) Choices are reviewed for each convention. Sometimes special resources for a particular presentation are included in the conference stores for a short time only.

Questions	AAAS Project 2061 1991-1994	Educational Products Information Exchange (EPIE) Institute	Eisenhower National Clearinghouse (ENC)
Purpose of Evaluation/Choice Process	database for educators	database of resources for educators	Information and database for educators
Type of organization	non-profit organization	non-profit organization	
What type resources does your organization examine?	science, math, technology print and non-print materials	marketed non-print products	Math and science program emphasis on federally funded materials
From what population do you select the resources for screening?.	published, public domain, print, non-print, human	electronic resources	All math and science materials donated to ENC
What process do you use to identify the population of materials?	literature, practitioners, conversation	published reviews, conversations	actively search for programs and information
Who is involved in the screening/evaluation process?	practitioner team, staff	4 staff	ENC staff
Do you differentiate among promising (P) and exemplary (E) resources ?	yes-student/classroom use effectiveness	no	Exemplary resources are chosen for special publications
What criteria are used to consider promising or exemplary? P=promising E=exemplary	P=meets criteria, has potential to be effective in learning environment E=effective in learning environment with testimony	Can it be used in a learning environment?	Collection Development Policy defines both P and E resources
What kind of evidence is needed to include a resource?	E=positive effect (anecdotal) in learning environment		All inclusive
Do you endorse a product or practice?	no-meets needs at this time	no-included in the database	No

[illegible]

Questions	International Society for Technology In Education (ISTE)	Instructional Resources California 1994 (rev)	NAS/Smithsonian National Science Resources Center (NSRC) •	NSTA
Purpose of Evaluation/Choice Process	reviews for journal	purchase of resources for California schools using state money	Resources for Teaching Elementary School Science -- published 1996	reviews published in journals
Type of organization	professional society	state	non-profit organization	professional organization
What type resources does your organization examine?	non-print resources	print and non-print resources	published science (print)	science only, no health, preK-12, U.S. published, print, non-print (only those currently in print)
From what population do you select the resources for screening?.		instructional content-related resources	science resources that support inquiry and hands-on learning (no text books or trade books)	real science with care to screen pseudo- or superficial science, no promotional materials
What process do you use to identify the population of materials?	submissions from publishers conversations	materials submitted by publishers	submissions from publishers (after letter of announcement)	1. 2 panels (print & non-print) 2. good reviews 3. generally useful material
Who is involved in the screening/evaluation process?	staff	state-trained educators for state review any educator in districts for local review	educators, scientists	1. 2-300 in areas of expertise nationwide 2. NSTA staff
Do you differentiate among promising (P) and exemplary (E) resources ?	reviews differentiate between P and E	no-ratings for any material are minimal, desirable or exemplary for accepted resource	yes-only exemplary are included in book	yes-exemplary only retained for resale, or review published
What criteria are used to consider promising or exemplary? P=promising E=exemplary	evaluation criteria designate P and E	even exemplary resources can present unequal results, so promising is not considered	evaluation criteria	E=meets criteria and reviewers' judgment
What kind of evidence is needed to include a resource?	good reviews	high correlation to listed guidelines	existing in resource, noted in evaluation	E=written review with details
Do you endorse a product or practice?	no	acceptance for purchase with state funds	included in Resources for Teaching Elementary School Science	choices sold in convention store, catalog

Questions	AAAS Project 2061 1991-1994	Educational Products Information Exchange (EPIE) Institute	Eisenhower National Clearinghouse (ENC)
How many people are involved screening each item?	one or more	staff	2 or more
What kind of record is kept of the screening/evaluations?--paper, computerized database, list only	evaluation form for database (print or electronic)	EPIE database	database
How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)	1. screen, 2. evaluate, 3. describe, 4.analyze, 5.record the data	accepted from good reviews elsewhere	published reviews and evaluations from other sources are included
How are decisions made (group, numeric average, discussion, combination)?	1. numeric based on evaluation form values 2. evidence	all-inclusive	
For how long is one review/acceptance valid? Are the accepted items revisited?	Each review is dated. revisited by others as time allows. Review every 3 years is planned		new reviews are added when available
Other items used for identifying/choosing materials		to make as much information as possible available to users	Science and math advisory boards

Questions	International Society for Technology in Education (ISTE)	Instructional Resources California 1994 (rev)	NAS/Smithsonian National Science Resources Center (NSRC)	NSTA
How many people are involved screening each item?	1 reviewer	state level-3-5 educators local- one or more	2 educators and 1 scientist	one or more and staff
What kind of record is kept of the screening/evaluations?--paper, computerized database, list only	paper	varies	print	written review
How is the evaluation conducted--(site visit, resource use in a classroom, stepwise process)	reviewing the material with use	state level-at least 2 individuals with group consensus	attend review training, then product reviewed by educators and scientist	materials distributed to reviewer
How are decisions made (group, numeric average, discussion, combination)?	individual	group discussion, leading to consensus	consensus among 2 groups, positive review in both pedagogy and content	1. good reviews 2. secondary review
For how long is one review/acceptance valid? Are the accepted items revisited?		state level- every 6 years	last edition published in 1988	choices reviewed for each convention
Other items used for identifying/choosing materials		guidelines are to identify exemplary resources. CA assumes different results in different environments		1. marketability for the season 2. relevant for limited use correlating to specific NSTA program/emphasis

311

312

RESOURCES

Questions	AAAS Project 2061 1991-1994		Educational Products Information Exchange (EPIE) Institute		Eisenhower National Clearinghouse (ENC)		International Society for Technology in Education (ISTE)	
Criteria used to evaluate programs	1. Evaluation form 2. Evidence form		1. All inclusive 2. Reviews		Inclusive			
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
quality	1. Screen=looks useful 4. Evaluation form, evidence			✓	✓		✓	
accuracy/currency	✓				✓		✓	
pedagogical effectiveness	✓ evidence				✓		✓	
correlation to research findings		✓			✓		✓	
correlation with standards or state frameworks	✓ evidence				✓			
equity, lack of bias	✓				✓		✓	
user appeal	✓						✓	
user involvement during use	✓				✓		✓	
ease-of-use	✓				✓		✓	
multiple content connections	✓				✓			
usefulness to others or in another setting		✓						
cost effectiveness	✓ item cost included							
Built-in user support	✓				✓		✓	
innovation or uniqueness		✓			✓			
Any other comments	1. Functions as it should				1. Evaluation from other sources included 2. Availability 3. Age appropriate		1. Technical quality 2. Technology is used effectively and appropriately	

Questions	Instructional Resources California 1994 (rev)	NAS/Smithsonian National Science Resources Center (NSRC)	NSTA
Criteria used to evaluate programs	Must correlate to state framework	Published in Resources for Teaching Elementary School Science	Content, illustrations, design, format
quality	✓ Instructional design	✓	✓
accuracy/currency	✓	✓	✓
pedagogical effectiveness		✓	
correlation to research findings	✓	✓	✓
correlation with standards or state frameworks	✓1st and essential criteria	✓	
equity, lack of bias	✓comply with state criteria	✓	✓
user appeal	✓	✓	✓
user involvement during use	✓	✓	
ease-of-use	✓	✓	
multiple content connections	✓	✓	✓no single-subject, comprehensive texts
usefulness to others or in another setting	✓		
cost effectiveness	✓		
Built-in user support	✓essential teacher's manual	✓	
innovation or uniqueness			
Any other comments	1. Utility programs evaluated separate from content 2. Highlights teacher/ learning variables encountered 3. Age appropriate 4. Supports declared objectives	1. Science/Society connection	1. Multiple points of view 2. Age appropriate 3. Materials designed to withstand use 4. Science/Society Connection

Appendix D

Summaries of Information About Evaluating Organizations

Appendix D
Summaries of Information About Evaluating Organizations

Programs

Annenberg
Concord Consortium
National Science Teachers Association (NSTA)
Northwest Regional Educational Laboratory (NWREL)
SouthEastern Regional Vision for Education (SERVE)

EVALUATION SUMMARY FOR MATH AND SCIENCE PROGRAMS

Organization:	The Annenberg/CPB Math and Science Project	
Address:	Corporation for Public Broadcasting 901 E Street N.W. Washington DC 20004	Toby Levine Communications, Inc. 7910 Woodmont Ave, Suite 919 Bethesda, MD 20814
Contact Person:	Scott Roberts (A/CPB)	Toby Levine (TLCI)
Telephone:	202-879-9711	301-907-6510
Fax:	*202-783-1036*	301-907-6916
e-mail:	mathsci@cpb.org	tk13@cornell.edu

Focus: A guide for educators on efforts that are stimulating fundamental and systemic changes in mathematics and science education in elementary and secondary schools. *The Guide* was developed to share information about the ways in which schools, districts, and states are approaching mathematics and science education, particularly in light of the publication of mathematics and science national standards, and to help interested individuals contact others with similar goals.

Product: A searchable database with more than 1000 entries outlining reform initiatives and organizations involved in math and science education reform. *The Guide* is available on disk or can be downloaded from the Annenberg/CPB WWW home page (http://www.learner.org/k12/).

Date of last evaluation: The fifth edition was released in August 1996; the sixth edition will be published in February 1997.

Approach: Federal, state, local, corporate, and privately funded initiatives that are involved with math and science education reform K-12 are documented in a searchable database for either MAC or MS-DOS/Windows platforms. Organizations dedicated to significantly improving K-12 math and science education are also included in the database. Most of the projects and organizations have at least some of the following characteristics:

- They are conducive to broad, systemwide changes in policy, operation, philosophy, methodologies, curricula, allocation of resources, and the ways in which teachers are introduced to new technologies and approaches.
- They are working to produce lasting change.
- They focus on methods of learning in which students construct their own understanding from inquiries and problem solving.
- They are aligning a state, district, or school with the national standards.
- They consider curricula that integrates mathematics and science and/or uses technology in innovative ways.

Difficulties encountered: Locating information about a large number of programs with a small staff, differentiating between the Eisenhower National Clearinghouse and *The Guide*, and disseminating disks to a maximum number of educators, especially those with the greatest need, presented a challenge.

Additional notation: Curriculum and instructional materials are typically not included. . Regular updates (with new entries and updated abstracts as programs progress) are published.

EVALUATION SUMMARY FOR MATH AND SCIENCE PROGRAMS

Organization: Concord Consortium
Address: 37 Thoreau Street
Concord, MA 01742
Contact Person: George Collison
Telephone: 508-369-4367
Fax: 508-371-0696
e-mail: george@concord.org

Focus: Mathematics and science projects/programs selected to be used as models in distance learning instruction.

Product: Projects/programs selected to be used as teaching components in distance-learning programs nationwide. The programs would be used to develop professional techniques for excellent distance-learning delivery and, at the same time, would be the courses available for student learning.

Date of last evaluation: 1996

Approach: The highly qualified selector sifted through courses applicable for grades 9-12 mathematics and science learning. In all cases, he searched for innovative materials and programs that could create a powerful impact on students. He looked for teachers who considered themselves scientists and teacher/researchers. The programs selected were required to have existing materials available for use, as well as some indication of transferability. Criteria such as a superb vision, commitment to equity, awareness of present student populations, thorough coverage of content, in-depth understanding of science standards, innovative strategies, inclusion of problem-solving, promotion of "in-culturization" (infusion of science into all aspects of everyday living) of science, and transferability guided the selection of programs. Older programs that could be modernized with a minimum of effort were also considered.

Difficulties encountered: Some promising programs were not far enough developed to consider. Because there is an understanding of diminishing funding and need in small schools the Concord Consortium is committed to developing good distance learning instructors. Teachers would receive training in the chosen exemplary courses to disseminate them more broadly.

Additional notation: This project was recently funded with a U. S. Department of Education Challenge Grant.

EVALUATION SUMMARY FOR MATH AND SCIENCE PROGRAMS

Organization: National Science Teachers Association (NSTA)
Address: 1840 Wilson Blvd.
Arlington, VA 22201-3000
Contact person: Phyllis Marcuccio
Telephone: 703-243-7100
Fax:
e-mail:

Focus: Identifying programs of excellence in areas of elementary school, middle school, high school, preservice, and non-school science instruction. Science supervision is also a topic that is considered.

Product: Three to 13 programs from state nominees in each considered field were chosen and spotlighted in print and in other ways. The criteria are offered as guidelines for excellence in science education.

Date of last evaluation: 1987

Approach: A network of 52 (the 50 states plus D.C. and Puerto Rico) state committees nominated programs in the areas of elementary school, middle school, earth science, physical science, biology, chemistry, physics, STS, environmental science, and energy science instruction and the areas of inquiry, science teaching, and career awareness, science in non-school settings, preservice teacher education, and science supervision. A national panel of judges drawn from professional and educational scientific organizations chose three to 13 programs in each field that most closely approached the national criteria for that field in the areas of goals, curriculum, instruction, evaluation, and teacher qualifications. In some cases evaluation areas were added to more thoroughly reflect science education needs. The list of nationally recognized programs, their teachers, and school locations are published by NSTA in *Criteria for Excellence*.

Difficulties encountered:

Additional notation: NSTA also reviews resources (see Resource summaries). This evaluation program is no longer active.

Appendix D
Summaries of Information About Evaluating Organizations

Resources

American Association for the Advancement of Science (AAAS)
Project 2061 (1) (2)
California State Department of Education
Educational Projects Information Exchange (EPIE) Institute
Eisenhower National Clearinghouse (ENC)
International Society for Technology in Education (ISTE)
Smithsonian Institution/National Academy of Sciences
National Science Resources Center (NSRC)
National Science Teachers Association (NSTA)

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES
AAAS Project 2061 (1)

Organization: Project 2061, American Association for the Advancement of Science (AAAS)

Address: 1333 H St. NW
Washington DC 20005

Contact Person: Carol Muscara

Telephone: 301-468-1160 (current phone for contact person)
Project 2061 phone ; 202-326-6400 (contact = Mary Koppal)

Fax: n/a

e-mail: CMUSCARA@EATS.COM MKoppal@AAAS.org

Focus: Educator evaluation, documentation, and dissemination of resource information that could be used to enhance science education reform

Product: A relational database of teacher-evaluated resources, including instructional materials, projects, community resources, human resources, and professional development resources

Date of last evaluation: December 1994

Approach: Teams of practitioners from six school district centers nationwide developed evaluation criteria for resources, correlated them to Project 2061 *Benchmarks for Science Literacy* and curricula, and entered them into a relational database designed by the practitioners and developed at Project 2061. An emphasis was placed on what teachers felt they needed to enhance student learning. The national database was available to educators in six districts nationwide. Evaluations followed a five-step process.

- Screening the resource for usefulness included a rapid screening by teachers, eliminating obviously inadequate resources.
- Evaluating the resource on the basis of a series of 20 criteria developed by K-12 teachers during a two-year period. These criteria reflect qualities educators felt were essential for exemplary resources. Two levels of evaluation were noted: (1) teacher opinion of potential effectiveness, and (2) teacher evaluation after use in educational setting.
- Describing the resource to provide maximum access and information for other educators. Descriptive information included access information, cost, content summary, subject synopsis, and grade relevance.
- Analyzing the resource to identify and document those Benchmarks that could be addressed to some degree using the resource. Both pedagogy and content evidence were considered for Benchmark correlation. Actual evidence within the resource was documented.
- Reporting the resource data included inputting all descriptive and evaluative information into a relational database. Data could be maintained as local (limited availability or interest) or sent to be included in a national database. The national database was updated regularly and shared with six pilot school districts nationwide.

EVALUATION SUMMARY FOR MATH AND SCIENCE PROGRAMS

Organization: Northwest Regional Educational Laboratory (NWREL)
101 SW Main Street, Suite 500
Portland, OR 97204

Contact Person: Rex Hagans

Telephone: 503-275-9594

Fax: 503-275-9584

e-mail: HAGANSR@NWREL.ORG

Focus: Consumer-oriented information about effective programs and resources for educators in the five state NWREL region

Product: Information is published in quarterly NWREL magazine, occasionally in special reports, and disseminated online and informally.

Date of last evaluation: Ongoing

- *Approach:* NWREL provides information about programs and resources for educators in its five-state (Alaska, Idaho, Oregon, Montana, Washington) region. Constituents want and need to know about effective practices within their region, but have neither the time nor the connections to do so. While there are no specific, formal criteria for recognizing a program or resource, several components are considered, including
    - specific alignment with math and/or science curricula and/or content standards
    - evaluation approach and data

Four categories of programs and resources are recognized. Information is made available to any educator in the NWREL region.

- Exemplary - as defined by the US Department of Education proposed guidelines
- Promising - as defined by the US Department of Education proposed guidelines
- State recognized - state funded programs or others that state educators or administrators say people ought to know about (some are officially recognized, others are not). These programs are not necessarily part of the National Diffusion Network (NDN), but might eventually be.
- Grass-roots successes - projects/programs that have gained the attention of peer educators and that show promise for enhancing teaching and/or learning.

Difficulties encountered: Disseminating information about effective programs and/or resources has been and continues to be difficult. Internet is a big step forward, but dependence on electronic communication overlooks those who are not connected. The focus is that educators want and need to know about good resources, not the whole world of resources.

Additional notation:

EVALUATION SUMMARY FOR MATH AND SCIENCE PROGRAMS

Organization: SouthEastern Regional Vision for Education (SERVE)
345 South Magnolia Drive, Suite D-23
Tallahassee, FL 32301-2950

Contact Person: Ms. Francena Cummings

Telephone: 800-854-0476

Fax: 904-671-6010

e-mail:

Focus: Evaluation and dissemination of promising practices nationwide

Product:

Print: Promising Practices in Mathematics and Science Education-1995: A Collection of Promising Educational Programs & Practices from the Eisenhower Mathematics and Science Regional Consortia--*available*: from ten Regional Consortia and Eisenhower National Clearinghouse

Includes full information for 50 programs chosen in 1995 and abstract information for 67 programs from 1994

Online: Searchable database of 1994 and 1995 selected mathematics and science programs--*available:* gopher.rbs.org

Date of last evaluation: 1995

Approach: SERVE takes the lead role to identify, evaluate, and validate promising practices submitted by ten Regional Consortia. The evaluation is a four-tiered process.

- First, the best practice applications are chosen by a panel of educators at each regional consortium. To be selected at the regional level, programs must be innovative, support emerging national mathematics and science standards, have evaluation data to demonstrate effectiveness, and be able to be transferred to other schools and/or educational settings. Programs meeting the criteria and presented most effectively on an extensive application form are submitted for national consideration.
- The national panel of representatives from each Regional Consortium chooses the best from those submitted by the Regional Consortia to provide the second level of evaluation.
- A Validation Team, with at least two members, visits each nominated site to provide the third level of selection. Site visits are used to address any concerns raised during the review process and to provide a first-hand validation of the nominated program. A numeric scoring for pedagogical features, innovation, effectiveness and transferability is the result of each visit.
- Final selections of programs to be recognized nationally are made based on the basis of national panel scoring and site-visit scoring.

Difficulties encountered: Even though site visits (validation process in step three above) are expensive and time consuming, they provide valuable first-hand confirmation for project written descriptions submitted by applicatants.

Additional notation: No commercial programs are considered

Difficulties encountered: Completing evaluations required teacher time and commitment. Because teacher/researchers were building, working with, and populating an emerging database, frustrating problems were sometimes encountered. Once the database was developed and stabilized, frustrations dissipated.

The most effective evaluation time was concentrated during the summer. Two weeks of consistent evaluation and analysis time were maximum amount of time for the teachers before efficiency began to decline.

For real effectiveness, all components - a functional database, adequate delivery hardware, current resource information, evaluations, curricula connections, *Benchmark* correlations, assessment tools, and technical support are needed simultaneously.

Additional notation: This is an early resource evaluation effort undertaken by Project 2061. Their most recent, in-progress effort (AAAS Project 2061 (2)) is described following this summary.

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES
AAAS Project 2061 (2)

Organization: Project 2061, American Association for the Advancement of Science (AAAS)
Contact Person : Jo Ellen Roseman
Telephone: 202-326-6752
Fax: 202-842-5196
e-mail: jroseman@aaas.org
Focus: To develop an analytical procedure educators can use to help them identify curriculum materials that support science literacy goals.
Product: CD-ROM and print editions of the guide, *Resources for Science Literacy* planned
Date of last evaluation: Work in progress
Approach: Working with teachers, teacher educators, and materials developers, Project 2061 has undertaken a year-long effort to design, pilot-test, and then field-test a procedure that evaluators the extent to which a curriculum material-ranging from a one-week unit to a multi-year textbook series-is likely to help students achieve the science literacy goals recommended in Project 2061's publications *Science for All Americans* 1989 and *Benchmarks for Science Literacy* (1993) and in the National Research Council's *National science Education Standards* (1996). The analysis can be used to: (1) improve decisions about the selection of materials, (2) identify shortcomings in existing materials and possibilities of improving them, and (3) inform the development of new materials. Analysis of how well materials promote science literacy involves considering both the specific content in the material and instructional strategies to support student learning of that content. The analysis work is organized into four steps.
Preliminary Evaluation: The goal of this step is a list of benchmarks and standards that seem to be worth pursuing in the subsequent analysis. Reviewers first search quickly through the material to make a preliminary list of all the specific learning goals that would seem likely to be targeted. The "material" examined includes both what the student sees and the instructional advice given to teachers, as in Teacher's Editions or accompanying guidebooks. The analysis is therefore of the full set of activities that its authors have explicitly prescribed for students. The material is then examined more carefully to locate and record all places where each learning goal is actually served - e.g., particular readings, experiments, discussion questions. Then, based on the number of sightings and the extent of the occurrence of each, benchmarks and standards are prioritized and those in the top category are pursued.
Content Analysis: The purpose here is to determine how well the specific content of the material matches the content of the "seemingly central" benchmarks and standards identified in the preliminary evaluation. Benchmarks (henceforth, to be taken as "benchmarks and standards") ranked highest are pursued through a more rigorous content analysis that involves giving more precise attention to both ends of the match in question. Once the goals themselves are clarified, a reinspection of the material focuses more closely on other criteria such as the specificity of the content match; level of

sophistication; whether the material addresses only a part of a benchmark, the whole benchmark, or multiple benchmarks, and whether the material contains text or activities that go beyond the science literacy goals.

Instructional Analysis The goal of this step is a rating and documentation of how consistent the material's treatment of each specific learning goal is with what is known about student learning. Ratings give readers a quick "bottom-line" view of how well specific benchmarks are addressed in the material. The documentation provides the evidence on which ratings are based.

Even if attention to the specific content of *Benchmarke* and *National Science Education Standards* (NSES) is necessary to promote science literacy, it is not sufficient. Principles of effective instruction that are based on research and/or on long experience are often not evident in curriculum materials. SFAA describes important characteristics for instruction in Chapter 13. Effective Learning and Teaching and the NSES emphasizes features of good teaching by creating independent teaching standards and by giving them a preeminent position in the document. Seven criteria have been identified to serve as a basis for the instructional analysis. Cast as questions, these are:

(1) Purposefulness: Does the student material explicitly identify the learning purposes of each activity and of the connected sequence of the activities it proposes? If not, does it prompt the teacher to make the purposes clear at the beginning and throughout? Does it cast the purposes in contexts of interest to students?

(2) Student Knowledge Base: Does the material inform teachers on what ideas students commonly hold with regard to the benchmarks of interest? Does it call upon teachers to find out what their own students believe about the ideas to be addressed, and then suggest how to build on those student conceptions? Does the material identify the knowledge and skills needed in order to participate successfully in the proposed activities?

(3) Active Engagement: Does the material include ;activities that provide first-hand experiences with phenomena relevant to the targeted benchmarks? When that is not feasible, does it make use of demonstrations, videos, pictures, models, simulations, and other surrogates for actual hands-on experience? Does the material place emphasis on having students see the question or problem as important to address? Does the material call for students to encounter an idea or skill in several different contexts?

(4) Scientific Ideas: Does the subject material have students develop an evidence-based argument for the benchmark ideas? Are the scientific ideas given more attention in the materials than technical terms? Does the material link ideas to emphasize that learning is not made up of isolated pieces of information? Does the material provide students with many opportunities to apply the ideas they are learning in order to emphasize their usefulness?

(5) Student Reflection: Does the material routinely include suggestions for having students express, clarify, justify, and represent their developing understandings? Are suggestions made for when and how students will obtain feedback from peers and teachers, and how teachers should respond to it? Does the material call upon

students to check their own progress, determine whether or not their thinking is changing, and if so, for what reasons?

(6) Assessment: If assessment items are included in the material, do they accurately reflect the substance of the targeted learning goals? Does the material include assessment tasks that place more emphasis on understanding and applying ideas than on memorizing terms, laws, and formulas. Are some assessment items embedded in the materials for teachers to use along the way for guidance in modifying planned activities?

(7) Learning Environment: Does the material contribute to the professional growth of teachers who use it in their classrooms? Does it help teachers create a classroom that holds high expectations for all students and that welcomes student curiosity, rewards creativity, encourages a spirit of healthy criticism, and avoids dogmatism?

Profile The analysis concludes with a report that summarizes the findings and draws some general conclusions about the usefulness and cost effectiveness of the material as a whole. Educators using the report as a basis for selecting materials will want to know both the material's benefits - i.e., what students might learn - and its costs - i.e., how much time the material requires. They might also want to be alerted to other important strengths and weaknesses of the material. Educators interested in improving the material will want more detailed information about the material's shortcomings.

Difficulties Encountered: Identifying curriculum materials that address less-traditional topics - the nature of science, historical perspectives, or systems, for example - and developing a strategy for evaluating them has proven to be problematic because (a) few materials exist that address these topics well and (b) educators are less knowledgeable about these topics. Also, the procedure takes more time and reviewers need a great deal more content knowledge and experience in applying the procedure than had been anticipated.

Additional notation: A 4-5 day training workshop in the use of the analysis procedure has been developed and field-tested. The training makes use of case studies to illustrate the application of the analysis criteria, gives participants opportunities to practice applying them, and provides feedback on their efforts.

This summary was written by AAAS Project 2061. An earlier resource evaluation effort is described in AAAS Project 2061 (1).

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES

Organization: California State Department of Education
Address: California Instructional Technology Clearinghouse
Stanislaus County Office of Education
801 County Center Three Court
Modesto, CA 95355
Contact Person : Harry E. Bakker
Telephone: 209-525-4993
Fax: 209-525-4984
e-mail: hbakker@stan-co.k12.ca.us
Focus: c
Criteria to be used statewide for evaluating instructional technology
Date of last evaluation: Ongoing

Approach: In California, the statewide adoption of materials for science and mathematics is a formal, rigorous process. Materials to be considered are submitted to the state by publishers. The State Department of Education, with the help of teams of educators from K-12 and university communities, develops criteria for evaluation. Teams of evaluator/educators are convened and familiarized with the criteria, their meaning and significance. Each evaluator is provided with materials to be reviewed. Evaluator teams must evaluate resources based on each criterion with careful annotation of evidence. Evaluators then are reconvened and must reach consensus on the rating of each criterion for a resource. Only those resources meeting strict state developed standards can be bought using state material funds. It is not required that any material be adopted for a given grade level.

In addition, more than 20 educators from K-12 and university communities recently revised 1991 guidelines to be used to select technology-based instructional materials for California schools. While this project deals with instructional technology, other state sponsored groups develop criteria for print and other non-print resources. The criteria emphasize five areas: curricular match, instructional design, content, interest (to learner), and technical quality. Several subcategories further define useful instructional technology materials in instructional design (program design, presentation design, interactivity, and teacher support), and in technical quality (audio, visual, program operation, and operation support. Each criterion is broken into specifications that designate three levels of materials: essential (must be present for resource to be used in CA), desirable (goes beyond the CA minimum) , and exemplary (goes beyond CA desirable criteria and can be highly recommended for use). The criteria are available to districts for choosing materials.

Difficulties encountered: Since California reviews content-related materials using a six-year cycle, newly developed materials must wait for up to six years to be adopted. The textbook/materials adoption process is a rigorous and expensive program. Publishers sometimes do not submit materials to the California adoption process to avoid being excluded from adoption.

Additional notation: Districts can use funds other than state funds to purchase materials that were not adopted.

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES

Organization: The Educational Products Information Exchange (EPIE) Institute
Contact Person: Ken Komoski
Address: 103-3 W. Montauk Hwy
Hampton Bays, NY 11946
Telephone: 516-728-9100
Fax: 516-728-9228
e-mail: EPIE INST@AOL.COM

Focus: Providing The Educational Software Selector (TESS), an online database of electronic products with educational value that is primarily accessible through the States Consortium for Improving Software Selection (SCISS). Schools in non SCISS states may purchase TESS on CD-ROM.

Product: An online database of resources for educators.

Date of last evaluation: 1996

Approach: The TESS database tends to be more inclusive than exclusive. Its goal is to help educators get better products for instructional use. A team of EPIE staff members use published reviews from more than 70 review sources to include materials in the database. This inclusion provides large numbers of resources for users to sift through using individual search-criteria to narrow the field of materials. A user-centered set of criteria is applied to resource choice for database inclusion. Generally, if there are no materials existing in the database to perform a particular function, a resource that can be stretched to fill that need will be included. The philosophy is that if a teacher might be able to use it creatively, it should be included. TESS is updated monthly and updated CD-ROMs are issued twice each year.

Difficulties encountered: During the past five years, only 1/3 to 1/2 of all resources nationwide have received published reviews. If one relies only on published reviews, some useful materials would be overlooked by educators.

Additional notation: The EPIE director proposes a universal software evaluation tool, *Software Multimedia Assessment and Reporting Tool (SMART)*. It will be research-based, user-friendly, WWW-based tool that educators and students will be able to use to submit evaluations to be shared via the TESS database.

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES

Organization: Eisenhower National Clearinghouse (ENC)
Address: 1921 Kenny Road
Columbus, Ohio 43210
Contact person: John Monk
Telephone: 614-292-6840 or 800-621-5785
Fax: 614-292-2066
e-mail: info@enc.org

Focus: To improve access to mathematics and science resources for the K-12 education community.

Product: ENC maintains a comprehensive collection and catalog of K-12 curriculum materials for K-12 educators (http://www. enc.org)

Date of last evaluation: Ongoing

Approach: ENC collects donated math and science resources including federally funded resources developed after September 1992, and it has created a database of information about those resources. More than 10,000 items are currently housed in the ENC library, 6000 of which have been input to the database. All resources are annotated in the database. Materials that are cutting-edge, align with national content standards, or of high quality are given priority for database inclusion. In addition, included resources must be thorough, comprehensive, and readily available. ENC locates and includes evaluative information about the resources from other sources (field tests, journals, awards). Each resource abstract is written and reviewed by content specialists. ENC does not evaluate resources, but rather gives priority to those considered of quality. It supports other national evaluation efforts by providing access to resources at two collection sites. The complete collection is housed at ENC headquarters at the Ohio State University. A more limited collection of materials is housed at the ENC Capital Collection located at George Washington University. A collection of virtual (electronic) resources (web sites, online documents) is available through ENC. These items, with complete access information are described in the ENC online catalog.

Difficulties encountered: Donated materials are the greatest proportion of items included in the ENC system, therefore, at any given time, some effective materials may not yet be in the ENC collection, in the ENC online catalog, or be overlooked. Also, ENC does not evaluate materials to select materials for inclusion in the ENC collection. Such evaluation would (1) restrict the flow of materials to the collection, (2) be unacceptable to some ENC clientele, and (3) complicate the acceptance policy and, therefore, limit the scope of the database.

Additional notation: The September 1994 *Collection Development Policy* published by ENC carefully documents the parameters used to develop the ENC database. This document is currently being revised and will be made available online when revisions are complete.

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES

Organization: International Society for Technology in Education (ISTE)
Address: 1787 Agate St.
Eugene, OR 97403
Contact Person : Anita Best
Telephone: 503-346-2400
Fax: 503-346-5890
e-mail: anita_best@ccmail.uoregon.edu
Focus: Educational software reviews that categorize resources for educators
Date of last evaluation: Ongoing

Approach: ISTE has defined a set of criteria to be used by field-based educators who evaluate electronic instructional materials for publication in ISTE's journal, *Learning and Leading with Technology.* The criteria are sorted into categories of instructional design, content, teacher support, and technical quality. In all cases, for resources to be considered exemplary, reviewers look for projects/resources where students learn by doing and where the technology allows students to manipulate data or use unique tools to construct learning. Publishers are given an opportunity to comment on reviews. Field-based evaluators submit reviews to ISTE that are edited before publication.

Historic Perspective: ISTE has been an educational technology advocate since the early 1980s.

Additional notation: ISTE will soon publish a book titled *ISTE Guidelines for Evaluating and Selecting Interactive Technology.*

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES

Organization: Smithsonian Institution/National Academy of Sciences (NAS) National Science Resources Center (NSRC)
Address: Smithsonian Institution, MRC 403
Arts & Industries Bldg., Rm. 1201
Washington, DC 20560
Contact Person: Evelyn Ernst
Telephone: 202-287-7247
Fax: 202-287-2070
e-mail: eernst@nas.edu
Focus: To provide annotated guides to curriculum materials and sources of information and assistance for teaching hands-on, inquiry-centered science in elementary and middle schools.
Product: The National Science Resources Center produced *Resources for Teaching Elementary School Science*, published early in 1996. The book annotates 350 curriculum materials; 115 teachers reference books and periodicals, and 450 facilities and organization that provide programs and materials for students and teachers.
Date of last evaluation: 1994-95 (elementary school guide) and 1996 (middle school guide)
Approach: To select curriculum materials for the elementary science resource guide, the NSRC convened panels of educators and scientists, providing them with the NSRC Evaluation Criteria for Curriculum Materials. The NSRC criteria are based on the national science education standards and reflect the work of the NSRC with educators nationwide since 1985. The criteria consider both pedagogy and content. The review process consisted of two phases. In the first phase, panels of teachers and science educators were convened on-site, where they received training and then spent up to 10 days reviewing materials that had been assembled by NSRC. Several educators independently evaluated each resource, using the following criteria:

- alignment with goals for elementary science teaching and learning defined in the National Science Education Standards
- focus on inquiry and activity-based learning
- incorporation of an effective instructional approach
- use of an effective format and design
- lack of bias

To select facilities and organizations for inclusion in the sections on sources of information and assistance in the guide, the NSRC used data collected in a form as survey. (No site visits were made.) The survey focused on programs, services, and materials provided in support of hands-on, inquiry-based elementary school science education. Only institutions that provided evidence of a significant effort to support the teaching of science in elementary schools are annotated.

Historic Perspective: NSRC solicited and reviewed resources to completely revise and update its 1988 publication for K-6 resources.
Additional notation: Computer software, audiovisual materials, science trade books, and elementary school science textbooks are not included in the guide. In all cases, expectations for core (comprehensive) resources were greater than those for supplementary materials.

EVALUATION SUMMARY FOR MATH AND SCIENCE RESOURCES

Organization: National Science Teachers Association (NSTA)
Address: 1840 Wilson Blvd.
Arlington, VA 22201-3000
Contact person: Phyllis Marcuccio
Telephone: 703-243-7100
Fax:
e-mail:

Focus: (1) NSTA chooses publications to be resold in its stores at headquarters and at national and regional conferences. (2) Resource reviews are published in its widely circulated journals.

Product: (1) Publications are reviewed and chosen for resale on the basis of good science and pedagogy content.

(2) New resources are reviewed by practitioners whose reviews are published in age appropriate association periodicals.

Date of last evaluation: ongoing

Approach: (1) A NSTA staff team reviews 500-600 education-oriented science resources annually. U.S.-published print and non-print resources such as student books, trade books, teachers guides, activity guides, kits, software, research findings, and "how-to" books are carefully reviewed, especially to eliminate pseudo-science, anthropomorphism, superficial science, and other non-science trends. Attention is given to focus topics at each convention, so the resources are constantly changing. Marketability is one of the criterion for choice. In some instances, special publications keyed to individual presenters or programs are included for a short time. Approximately 40% of the resources chosen are NSTA publications and 60% are from other sources (2) Evaluator-practitioners regularly review newly published (copyright must be the present year +/- one year) K-12 resources such as books, kits, audio-visual materials, software, and games that deal with science, science teaching, and science learning. Two to three hundred practitioners with content expertise are sent resources to evaluate on the basis of content, pedagogical, design, and usefulness criteria. Some high school mathematics resources are considered. Health-education materials are not considered. No comprehensive texts are reviewed.

Difficulties encountered:

Additional notation: NSTA also publishes *Criteria for Excellence*, criteria that define and describe exemplary science programs in the United States. (see Programs/Practices summaries).

Appendix E

Sample Evaluation Criteria From Organizations

Appendix E
Sample Evaluation Criteria from Organizations

Programs

From
National Science Teachers Association (NSTA)

An Introduction to the Criteria

he Search for Excellence in Science Education was inaugurated in 1982 to arry out the National Science Foundation's attempt to define and to escribe exemplary science programs in the United States. The Search rocess, based on NFS's Project Synthesis, brought together visions of the deal in science education from experts in a wide range of scientific isciplines, synthesized those visions into criteria for excellence, and pplied them to actual programs identified during national searches across he United States.

The experts were drawn from a wide range of professional and ducational scientific organizations, ranging from NSTA, the parent of the Search, through the National Science Foundation and the National Assessment of Educational Progress (partners in Project Synthesis), to the American Association of Physics Teachers, the American Chemical Society, the National Association of Geology Teachers, and the National Association of Biology Teachers, as well as the divisions of NSTA and its division affiliates, the National Science Supervisors Association, the Council of Secondary Science Supervisors, the Society for College Science Teaching, the Association for the Education of Teachers in Science, the National Association for Research in Science Teaching, and the Council for Elementary Science International. The leaders of the program continued to involve their peers at all levels of education and in all science disciplines.

Project Synthesis identified four kinds of goals for today's science learning: personal, societal, vocational, and academic. Academic success, the most traditional and most narrowly conceived objective of science education, was the one area in which the members of the project saw satisfactory achievement in 1981. But all citizens of the 90s need science competence to make the most of their individual potential, and to cope with an increasingly technical world. And as participants in a democratic society, individuals must be prepared to make informed, responsible decisions about science-related social issues. Moreover, as the pace of change accelerates, new careers are being created almost daily, as diverse as the interests and aptitudes of students themselves. Alerting students to the ever-widening range of opportunities in careers related to science and technology, teaching science for life, and teaching science for citizenship: These are the three challenges the Project posed to the science programs of yesterday, today, and tomorrow.

The Search for Excellence sought out and honored programs in one or several fields each year. Categories included general approaches for specific age groups, such as middle/junior or K-6; standard academic disciplines, such as biology and chemistry; and interdisciplinary areas, such as energy education, science as inquiry, and science/technology/society. In each area the Search identified hallmarks of excellence in terms of goals, curriculum, instruction, evaluation, and teacher qualifications.

After the criteria had been identified, the search began. Each year a network of 52 state committees (one for each state plus the District of Columbia and Puerto Rico) sought out programs that approached the national criteria in each target area. From the states' nominees, a national panel of judges selected 3 to 13 programs in each field that most closely approximated the ideal.

No real life program ever met all the criteria, or ever will; but after all the hard work of analysis, synthesis, and search, it is possible to get a glimpse of the future of science education. The criteria are just the rudiments of that future. Its real articulation—and the living definition of the excellence the program sought—is to be found in the classrooms of our exemplars, the people whose commitment to excellence made our search rewarding.

The Search for Excellence builds for the future by spotlighting shareable models—exemplary programs that really work—and recognizing the talent, dedication, and generosity of the exceptional individuals who make them work. The future is up to you, the readers of this book.

Look at your program in the light of these criteria. Congratulate yoursel and your colleagues on your program's strengths. Then plan together to build towards these ideals in every area where your program can be improved.

Robert Yage
NSTA President, 1982-8.

Appendix E
Sample Evaluation Criteria from Organizations

Programs

From
SouthEastern Regional Vision for Education (SERVE)

SERVE

SouthEastern Regional Vision for Education

Promising Practices in Mathematics and Science Education

Application

For more information concerning the Promising Practices Program, contact your Eisenhower Regional Mathematics and Science Consortium:

Anne Tebo, Projects Coordinator
SERVE Consortium for Mathematics and Science Education
345 South Magnolia Drive, Suite D-23
Tallahassee, Florida 32301-2950
(800) 854-0476

Applications must have an abstract and superintendent/designee's signature, and be postmarked by December 20, 1993, in order to be considered for 1994.

TABLE OF CONTENTS

Page

APPLICATION PROCEDURES iv

PART ONE: CONTACT INFORMATION 1

PART TWO: DESCRIPTION 2

PART THREE: PEDAGOGICAL FEATURES 5

PART FOUR: HIGHLIGHTS 9

PART FIVE: EFFECTIVENESS 11

PART SIX: TRANSFERABILITY 12

PART SEVEN: ABSTRACT 14

APPLICATION PROCEDURES

1. **PURPOSES:** The purposes of the Promising Practices in Mathematics and Science Education project are to:

   - identify promising curriculum materials, instructional strategies, and assessment tools , and
   - disseminate information about these practices/programs nationally.

   Information about Promising Practices is disseminated nationally through a publication, diskette, and on-line database via Eisenhower Regional Consortia for Mathematics and Science Education and the Eisenhower National Clearinghouse. The publication (hard copy or disk) and database describe each selected program/practice. This brings attention and recognition to the developers. (Note: This may lead to many requests for information.)

2. **ELIGIBLE PROGRAMS:** Eligible programs can range from individual classroom to systemwide efforts. Commercially marketed programs are not eligible. However, those developed or distributed by non-profit institutions are eligible along with those distributed by for-profit organizations for free or at cost. The Consortia are seeking programs that are the following:

   - innovative
   - support emerging national mathematics and science standards
   - have evaluation data to demonstrate effectiveness
   - can be transferred to other schools and/or educational settings

3. **APPLICATION SUBMISSION:** The original application and 3 copies must be postmarked **December 20, 1993.** In order to be processed, an application must be signed by the superintendent or appropriate official.

4. **REVIEW PROCESS:** A panel in each region will evaluate and select programs and practices which meet the criteria mentioned above. The second stage of evaluation involves a national review of each region's selections. The final stage is a site visit. (Selected programs/practices are expected to be available for visitation and to provide additional information about their program/practice.)

PART ONE: CONTACT INFORMATION

CONTACT INFORMATION: Please provide the following information regarding your program/practice. If there is more than one contact person, please provide contact information on a separate sheet that is clearly marked and labeled. The signature of the superintendent or designee is required for nominations of school-based programs/practices; for others, the signature of the project director is sufficient.

Project Name: ______________________________

Principal/Director: ______________________________

Contact Person: ____________________ Title: ______________

School Name: ______________________________

Address: ____________________ District: ______________

City/State: ____________________ Zip: ______________

Telephone: __________ Fax: __________ Home: __________

Internet e-mail: ______________________________

Superintendent/Appropriate Official: ______________________________
(Please print or type.)

______________________________ Signature

______________ Date

FOR OFFICE USE:

DATE RECEIVED: ______________ **REFERENCE CODE:** ______________

BEST COPY AVAILABLE

PART TWO: DESCRIPTION

Five categories are provided to assist you in a general description of your program/ practice: Date Initiated, Grade Levels, Subjects, Target Audience and Funding Sources

1. When was this program/practice initiated? ______________________________

2. This program/practice is designed specifically for which grade levels (check all that apply):

☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6
☐ 7
☐ 8
☐ 9
☐ 10
☐ 11
☐ 12
☐ Preschool
☐ Kindergarten
☐ Elementary grades K-4
☐ Middle grades 5-8
☐ Junior High grades 7-9
☐ Secondary grades 9-12
☐ Post-secondary

3. **SUBJECTS:** The major subject areas of your program/practice are (check all mathematics and/or science areas that apply):

MATHEMATICS EDUCATION

☐ Elementary School Mathematics (K-8)
☐ Secondary School Mathematics (9-12)
☐ College Mathematics
☐ Algebra
☐ Arithmetic
☐ Calculus
☐ General Mathematics (Introductory or remedial mathematics emphasizing everyday computational operations and practical applications, commonly intended for noncollege-bound secondary school students)
☐ Geometry
☐ Mathematical Applications
☐ Mathematical Concepts
☐ Statistics
☐ Technical Mathematics (Mathematics needed in technical occupations such as electronics)
☐ Trigonometry
☐ Other

____

SCIENCE EDUCATION

☐ Elementary School Science (K-8)
☐ Secondary School Science (9-12)
☐ College Science
☐ Astronomy
☐ Biological Sciences (Life Sciences subjects such as Ecology, Wildlife, etc.)
☐ Biology
☐ Chemistry
☐ Earth Science
☐ Environmental Education
☐ General Science
☐ Geology
☐ Integrated Science
☐ Marine Education
☐ Natural Sciences
☐ Physical Sciences
☐ Physics
☐ Science & Society
☐ Space Sciences
☐ Technological Literacy
☐ Other

____

____

____

4. **TARGET AUDIENCE:** This program/practice was designed particularly to serve the following target audience(s):

Students

- ☐ Adult Education Students
- ☐ Advanced Placement Students
- ☐ Culturally Diverse
- ☐ Economically Disadvantaged
- ☐ Emotionally Disabled
- ☐ English as a Second Language (ESL)
- ☐ Female
- ☐ Gifted/Honors
- ☐ Hearing-Impaired
- ☐ Learning Disabled
- ☐ Minority
    - ☐ African American
    - ☐ Asian American
    - ☐ Hispanic American
    - ☐ Native American
    - ☐ Pacific Islanders
- ☐ Physically Disabled
- ☐ Remedial
- ☐ Rural
- ☐ Suburban
- ☐ Urban
- ☐ Other ______________________

Other Target Audiences

- ☐ Administrators
- ☐ Community
- ☐ Elementary Teachers
- ☐ Guidance Counselors
- ☐ Parents
- ☐ Secondary Teachers
- ☐ Student Teachers
- ☐ Supervisors
- ☐ Other ______________

5. **FUNDING SOURCES:** Which of the following sources contributed to the development and implementation of your program/practice? (Check all that apply.)

- ☐ Chapter 1
- ☐ State/District Eisenhower Mathematics & Science Program
- ☐ National Eisenhower Mathematics & Science Program
- ☐ Other ______________________
- ☐ Chapter 2
- ☐ National Science Foundation
- ☐ School District Funds
- ☐ State Funds

PART THREE: PEDAGOGICAL FEATURES

Four categories are provided to assist you in describing the educational features of your program/practice: Educational Materials, Instructional Methods, Assessment, and Standards. Within the first 3 categories, please select the features which best describe your program/practice.

1. **EDUCATIONAL MATERIALS** (Check up to 5.)

- ☐ Educational Games
- ☐ Enrichment Materials
- ☐ Learning Modules
- ☐ Manipulative Materials
- ☐ Multicultural Materials
- ☐ Parent Materials
- ☐ Remedial Materials
- ☐ Resource Materials
- ☐ Technology-Based Materials
- ☐ Workshop Materials
- ☐ Other ________________

2. **INSTRUCTIONAL METHODS** (Check all that apply.)

- ☐ Cooperative Learning
- ☐ Experiential Learning
- ☐ Hands-on Learning
- ☐ Individualized Instruction
- ☐ Interdisciplinary Approach
- ☐ Journal Keeping
- ☐ Multicultural Approach
- ☐ Multimedia Instruction
- ☐ Peer Tutoring
- ☐ Problem Solving Approach
- ☐ Student-Centered Instruction
- ☐ Technology-Based Strategies
- ☐ Thematic Approach
- ☐ Other ___________________________

3. STUDENT ASSESSMENT (Check up to 5.)

- ☐ Alternative Assessment
- ☐ Criterion-Referenced Achievement Tests
- ☐ Holistic Evaluation
- ☐ Informal Assessment
- ☐ Interviews
- ☐ Norm-Referenced Achievement Tests
- ☐ Performance-Based Assessment
- ☐ Portfolios
- ☐ State Proficiency Tests
- ☐ Student (self) Evaluation
- ☐ Student Journals
- ☐ Other ______________________

4. **STANDARDS**

List up to 4 mathematics and/or up to 4 science standards your program/practice addresses particularly well. You may use the abbreviated list on page 8 as a guide. Feel free to use standards that are not listed.

MATHEMATICS:

1. ______________________________
2. ______________________________
3. ______________________________
4. ______________________________

SCIENCE:

1. ______________________________
2. ______________________________
3. ______________________________
4. ______________________________

Describe how your program/practice addresses 2 of the standards that you listed.

SELECTED NCTM MATHEMATICS STANDARDS*

This program/practice is designed to:

- use sound and significant mathematics
- build on students' prior experience and knowledge
- develop mathematics thinking skills that convince students of the validity of particular representations, solutions, conjectures, and answers
- engage students' intellect; pose questions and tasks that elicit, engage, and challenge each student's thinking
- develop students' mathematical knowledge and skills
- stimulate students to make connections and develop a coherent framework for mathematical ideas
- call for problem formulation, problem solving, and mathematical reasoning promote the development of all students' dispositions to do mathematics
- develop an instructional model based on the range of ways diverse students learn mathematics
- pose tasks based on sound and significant mathematics

* *National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA.*

SAMPLE SCIENCE STANDARDS

This program/practice is designed to:

- be accessible to all students
- build on students' prior experience and knowledge
- use an instructional model based on the scientific process such as: question, discover, create, communicate, and pursue new questions
- select science concepts that are developmentally appropriate, with illustrative examples drawn from the content of multiple disciplines of science
- develop scientific thinking skills such as drawing conclusions based on evidence, using inference, creating models
- develop scientific habits of mind such as curiosity, skepticism, honesty, living with ambiguity
- use authentic assessments to chart teaching and learning
- shift the role of teacher from imparter of knowledge to designer and facilitator of learning
- seek to find relevant and significant applications of science content and concepts to students' personal and community life
- show the connections among different scientific concepts and/or disciplines

PART FOUR: HIGHLIGHTS

Please answer the following questions to highlight the significant features of your program/practice. Limit your answers to the space provided.

1. What are the 2 or 3 key features of your program/practice?

2. What needs prompted the development of your program/practice?

3. What is the main goal of your program/practice?

4. What are the objectives that enable the program/practice to reach that goal?

5. What are the innovative features of your program?

PART FIVE: EFFECTIVENESS

Describe the success or effectiveness of your program/practice by answering the following questions. You may supplement this with any available evidence (e.g., test scores, achievement gains, survey results, increased enrollments, etc.)

1. What evidence or indicators did you use to determine the effectiveness of your program/practice? (e.g., absentee rate, parent comments, test scores, student attitude, portfolios, discipline, problems, etc.)

2. What were the data or results obtained from the evidence/indicators cited above?

3. How were the results used to improve your program/practice?

4. Has your program/practice received previous validation or recognition? ________
Please specify:

Year	Source/Description
________	________
________	________

PART SIX: TRANSFERABILITY

Six prompts are provided to assist you in documenting the transferability of your program/practice.

1. Is this program/practice being used at any other sites? If so, how many? Please list up to three other sites (school, city, state) at which the program/practice has been effectively implemented.

2. Describe the MATERIALS essential to replicate your program/practice:

   Where can these be obtained?

   Provide estimated costs for the above:

3. Describe the EQUIPMENT essential to replicate your program/practice:

   Where can these be obtained?

   Provide estimated costs for the above:

4. Describe the HUMAN RESOURCES or training essential to replicate your program/practice.

Where can these be obtained?

Provide estimated costs:

5. Are there other resources essential to replicate your program/practice? Please describe.

Where can these be obtained?

Provide estimated costs:

6. What existing support resources are there for educators using this program/practice (e.g., 800 number, newsletter, users network, users bulletin board, etc.)?

PART SEVEN: ABSTRACT

In the space below, describe your program/practice. The single-space abstract (300 words or less) should clearly explain the program/practice, who it serves, its effectiveness, and what resources are needed to support it. This abstract will be used to describe your program in the *Promising Practices* publication and diskette.

PROGRAM REVIEW AND SELECTION GUIDELINES FOR SERVE

As directed by the provisions contained in Rules and Regulations language of August 14, 1987, panelists evaluate submissions and award points on the basis of the following categories:

Voting Categories

Results (0–50 points)

Panelists determine the extent to which the results indicate that:

- the program, product or practice's effect is convincing relative to similar programs; and
- the outcome claims of the program, product or practices are valid.

Evaluation Design (0–40 points)

Panelists determine the extent to which the evaluation design:

- is appropriate for the program, product or practice;
- is based on a correct interpretation of relevant research and literature;
- demonstrates that a clear and attributable connection exists between the evidence of an educational effect and the program treatment; and
- accounts for rival hypotheses that might explain effects.

Replication (0-10 points)

Panelists determine the extent to which the program, product or practice can be used at other sites with the likelihood of achieving similar results.

After the panelists complete their reviews, they each rate the submission in the above categories. The scores for each category are then added together for an overall rating between 0 and 100. The scores of six panelists are averaged for a final total rating.

Program effectiveness review approval is granted if the average panel rating for the *Results* category is at least 40 points, and the average total rating is at least 70 points. If the mail review results in a total average rating between 50 and 69 points, the Chair reviews the panel members' written comments to determine whether the vote represents a clear disapproval or whether further review by an in-person panel is warranted.

A second review (in-person panel) is justified if the panelists' written comments indicate a need for further clarity about the project's design or evaluation evidence. The composition of the in-person panel might not be identical to the original mail review panel. Typically, panels are convened in Washington, D.C., when there are at least three eligible programs for review. When an in-person panel is convened, the project developer is invited to attend the review to answer questions.

Additional claim types will be developed as other types of programs seek PEP approval.

Claim Type 1 — Academic Achievement: Changes in Knowledge and Skills

Traditionally, programs claiming to result in greater knowledge or increased learning of skills have been the most likely to come before the panel. The claim may demonstrate gains in knowledge or skills by any type of learner—students at any grade level, teachers, or other adult learners.

Projects for which this model is most appropriate are instructional interventions that teach content or skills, or provide opportunities for students to apply knowledge. Examples are traditional school curriculum areas such as reading or mathematics and emerging subjects such as computer science and thinking skills, as well as areas such as adult literacy. Claims in this area are based on the observation of measurable changes in the target population.

Examples of Claims

- Acquisition of factual knowledge: Students in the physics project at three typical high schools made greater gains than the national norm group on a standardized test of physics knowledge.
- Acquisition of new types of knowledge (*i.e.*, knowledge not presented in a typical curriculum): When compared with a control group, students in a computer literacy course scored significantly better on reliable (split-half r = .93), locally developed tests of computer knowledge.
- Rapid acquisition of knowledge (*i.e.*, changes in the efficiency of learning): Students completing a 1-semester math course performed as well on a standardized test as did a matched comparison group of students taking the traditional 1-year course.
- Application of knowledge: In addition to making greater-than-expected gains in library reference skills on a standardized test, program students required significantly less assistance with research activities than comparison students, as measured by structured observations in school libraries.
- Acquisition of skills: Quantitative studies in eight separate sites, using various nationally known measures, showed significant advantages in the area of reasoning ability for students in a philosophy program over comparison group students.
- Application of skills: Project students achieved significantly better ratings on analytically scored writing samples than did comparison students in the regular language arts program.

Projects offering this type of claim often present evidence based on familiar measures. Chief among these are written tests, including standardized norm-referenced tests, locally developed tests, and criterion-referenced tests. Generally speaking, tests have the advantage that their reliability and validity can be determined using established psychometric methods.

Potential Panel Questions About Evaluation Design:

- Is this the strongest and most appropriate research design that could be undertaken given the nature of the project treatment, setting, and participants? If not, what are the reasons for not choosing another design?
- Have the inherent assumptions of the design been taken into consideration?
- Can the appropriateness of the comparison standard be demonstrated?
- How was the comparison group chosen? Is there evidence that it is similar to the project group in educationally relevant ways?
- If participants were selected on the basis of test scores, has a separate pretest been used in order to avoid the regression effect error? Have other measures been taken to counter the impact of regression?
- If a sample of program participants is used, is it a representative sample and has the sample been selected in a nonbiased fashion? Is the same true for the comparison group?
- Is the size of the evaluation sample large enough to generalize with sufficient confidence to the target population as a whole?
- Have sufficient numbers of learners remained in the study during the treatment period? Have the reasons for attrition and its effects been investigated?
- Have participants been selected in accordance with rules for the evaluation design?
- Is the timing of data collection appropriate and logical for the treatment and for the instruments used?

Instruments, Procedures, and Data Collection

The actual methods used to measure the changes produced by the project are of great importance, since no amount of analysis or argument can redeem a body of evidence that is flawed by improper choice of instruments, incorrect procedures, or contaminated data.

The panel needs sufficient information showing that the instruments are appropriate, reliable, and valid as measures of the project's claims. The less well-known the instrument, the greater the burden for the evaluator to establish these points. Whatever the instruments, it is important to indicate that they were administered in the proper way for both treatment and comparison groups. Possible sources of contamination in the data should be guarded against; if they are unavoidable, their effects should be acknowledged and, if possible, estimated.

Again, the less straightforward the data collection procedures, the greater the burden for the evaluator to document that procedures are credible. For example, if writing samples are used to demonstrate improvement in composition skills, they should be typed so that raters will not be influenced by penmanship or extraneous appearance factors, and pre- and post-writing samples should not be identified as such to avoid creating differential expectations of the two samples. All scoring should be done after the post-data is collected. Pre- and post-writing samples should be

Promising Practices in Mathematics and Science Education
Site Visit Validation Report

Purpose

The site visit is for verification of the accuracy of the descriptions in the nomination form that has been submitted by the district or school and previously rated by a review team of content area experts. In addition, the site visit will serve to address the concerns and issues raised about the program during the review process.

The Team

The Validation Team consists of at least two persons who may include a representative from the Eisenhower Regional Mathematics and Science Consortium and personnel from the Department of Education, a local school district, a university, or a business.

The Length of the Visit

Because the complexity and characteristics of programs differ, the length of visits varies. Generally, however, the team should be on-site no less than half a day and nor more than one day.

Team Report

The **Validation Team Report** consists of the attached cover sheet and issues/observations page.

Cover Sheet

All information on the cover sheet should be completed. The cover sheet should be signed by the reviewers.

Promising Practices in Mathematics and Science Education
Site Visit Validation Report

Title of Program ______________________________

Issues	Observations
Pedagogical Features 1. 2. 3.	
Innovation and Significance 1. 2. 3.	
Effectiveness 1. 2. 3.	
Transferability 1. 2.	

Outreach Capability

1. Are program developers able to handle telephone calls about the program? ________
2. Are program developers able to receive site visitors? ____________________
3. Are program developers able to provide materials for potential adopters? ________

CHECK LIST FOR SITE VISIT

Title of Program: *Problem Solving Approach* *MS004*

The following criteria has been used to recommend this program for Exemplary Successful practice. As a result of your site visit, please verify the information below by rating each task from strongly agree to strongly disagree.

PROGRAM DESCRIPTION:	SA	A	D	SD
1. The program/practice is consistent with the standards. COMMENTS:	1	2	3	4
2. The needs of the target population are served appropriately. COMMENTS:	1	2	3	4
3. The purpose and goals are appropriate for this program/practice. COMMENTS:	1	2	3	4
PROGRAM EFFECTIVENESS:				
4. The success of the program/practice is based on achievement of the stated purpose and objectives. COMMENTS:	1	2	3	4
5. The methods used to gather data and analyze results are appropriate for the program/practice and its objectives. COMMENTS:	1	2	3	4
6. There is strong and convincing evidence that significant change has occurred. COMMENTS:	1	2	3	4

Appendix E
Sample Evaluation Criteria from Organizations

Resources

From
International Society for Technology in Education (ISTE)

Draft of ISTE Guidelines for Evaluating and Selecting Interactive Technology Resources

Instructional Design

- The program is pedagogically sound
- Current educational research is embodied in and effectively used throughout the program
- The program promotes creativity, problem solving, and the development of higher order thinking skills in students
- The program makes effective use of interactive strategies
- Learners find operation of the program to be intuitive, with simple commands that seem to be transparent
- The program has a well-designed and appropriate hardware interface
- The presentation design enhances the learning experience
- Features in the program support the learning/teaching process
- The program allows students and teachers to create individualized instruction
- Use of multimedia enhances the learning experience
- The interest level is well suited to the learner
- Instructional tools are designed for ease of use and for meeting a variety of learner needs

Content

- The program content is presented impartially and without bias or distortion
- The program content is appropriate to student needs, curriculum area, purpose, and grade level
- The program content and design meet the needs of students at varied levels of English language acquisition
- The program has current, thorough, and relevant information
- Reference content on electronic media is fully and accurately indexed
- Search strategy software for electronic reference tools is designed to stimulate student research and to facilitate student access to information
- Search results can be displayed and printed appropriately
- Search strategies are adapted as appropriate for periodical indexes

Teacher Support

- Teacher support materials enhance the value of the program
- Helpful and complete instructions for using the program are provided
- Teacher support materials contain essential program information for planning lessons
- Teacher support materials contain suggestions for curriculum integration and for using the program in a variety of instructional settings
- Packaging is appropriate and provides essential operation information
- The publisher/producer/distributor provides good support.

Technical Quality

- The program and all components are reliable under normal use
- Program installation requires a minimal level of computer expertise
- High quality audio contributes to program effectiveness
- High quality visuals contribute to program effectiveness
- The program operates effectively in a network environment.

Appendix E
Sample Evaluation Criteria from Organizations

Resources

From
National Science Resources Center (NSRC)

In Press
National Academy Press
1996

DRAFT

PART 1

Few decisions have greater impact on the effectiveness of science teaching in the nation's schools than the process of selecting instructional materials. This selection determines to a large extent what will or will not be taught to children; it establishes the basis of teachers' professional growth opportunities in science instruction; and it accounts for major budget outlays for school systems.

Yet it is difficult for entire school districts, let alone individual classroom teachers, to find the time and resources to research the ever-growing volume of available curriculum materials, to assess them for scientific content and processes, and to arrive at the combination of materials suitable for their needs. Schools and teachers need authoritative information that addresses the educational and scientific aspects of teaching elementary school science to help make their selections.

In response to this need, the National Science Resources Center (NSRC), sponsored jointly by the National Academy of Sciences and the Smithsonian Institution, has produced *Resources for Teaching Elementary School Science*—an annotated guide to hands-on, inquiry-centered curriculum materials and sources of information and assistance for teaching elementary school science. This new volume is a completely revised and updated edition of the NSRC's best-selling resource guide, *Science for Children: Resources for Teachers*. The new edition focuses on curriculum materials published between 1985 and 1995 for kindergarten through sixth grade and on sources of information relevant to teaching science in the same grades.

The goal of the National Science Resources Center in developing *Resources for Teaching Elementary School Science* is to help teachers teach science more effectively. Thus, the NSRC has brought together in one source a

DRAFT

list of carefully reviewed and selected materials and resources. These curriculum materials and other resources support inquiry-based science teaching that fosters understanding of science concepts through hands-on student investigations. Teachers, principals, administrators in schools and school districts, science curriculum specialists, parents, and those involved in systemic reform of science education will find the guide a rich source of current information.

The materials and resources listed can be used to improve an existing program or to design a complete curriculum. It should be emphasized, however, that the guide is not a recipe for an elementary school science program.

Contents of the Guide

Following is a brief description of the contents and organization of the volume. It contains four parts:

- Part 1: Introduction to the Guide
- Part 2: Elementary School Science Curriculum Materials
- Part 3: Teacher's References
- Part 4: Ancillary Resources for Elementary Science Teachers

Part 2 contains about 350 individual entries that list and annotate curriculum materials. (The process by which these materials were selected is described below, in the section on "NSRC's Curriculum Evaluation Criteria and Review Process.") The overview in part 2 is followed by four chapters: chapter 1, "Life Science"; chapter 2, "Earth Science"; chapter 3, "Physical Science"; and chapter 4, "Multidisciplinary and Applied Science."

The annotations in these chapters are subdivided in the following categories: Core Materials, Supplementary Materials, and Science Activity Books. (The categories are defined in the part 2 overview.)

Chapter 5, "Curriculum Projects Past and Present," completes part 2, with information on major funded projects in hands-on elementary science over the years dating back to the late 1960s and early 1970s.

Part 3, "Teacher's References," has an overview and three short chapters of annotations: chapter 6, "Books on Teaching Science"; chapter 7, "Science Book Lists and Resource Guides"; and chapter 8, "Periodicals." Chapter 6 is an annotated list of about 50 volumes that provide background information and a broad range of pedagogical resources for good science teaching. Chapter 7 annotates about 25 directories and guides, including guides to science trade books for children and to materials and other resources. Chapter 8 annotates about 35 periodicals, including some magazines for children. The periodicals in the chapter were selected for their excellence as instructional tools, for the high quality of their articles and stories on scientific topics, for their appeal to children, and for their adaptability to classroom use.

Part 4 of the guide—"Ancillary Resources for Elementary Science Teachers"—contains two chapters that focus on facilities, associations, and federal and other organizations that have programs, services, and materials relevant to some aspect of hands-on, inquiry-based elementary school science education. The resources included in chapters 9 and 10 can significantly enhance the effectiveness of science education efforts.

Chapter 9, "Museums and Other Places to Visit," identifies almost 600 facilities—for example, museums, zoos, science and technology centers, and children's museums—to which elementary science teachers can take their classes for hands-on science experiences beyond the classroom. Annotations are provided for about half of those institutions—those considered to be making a significant effort to help teachers teach science more effectively.

Chapter 10, "Professional Associations and U.S. Government Organizations," presents annotated entries for about 120 institutions with a wide range of scientific, educational, and professional missions. The purpose of the chapter is to guide teachers to private and public sources of information, materials, and services that support elementary school science both directly and indirectly, and to identify science education facilities and relevant programs administered by U.S. government organizations.

Finally, the appendixes in the volume include a list of "Publishers and Suppliers" (appendix A) for curriculum materials and other publications annotated in the guide. Appendix B discusses and reproduces the NSRC evaluation criteria formulated for use in the review of curriculum materials.

Multiple indexes are provided to help readers access information quickly and efficiently.

NSRC's Curriculum Evaluation Criteria and Review Process

Consistent with the NSRC's philosophy of science teaching and with the recently published *National Science Education Standards* of the National Research Council, the materials included in this guide are hands-on and inquiry-centered. Briefly described, such materials provide opportunities for children to learn through direct observation and experimentation; they engage students in experiences not simply to confirm the "right" answer but to investigate the nature of things and to arrive at explanations that are scientifically correct and satisfying to children; they offer students opportunities to experiment productively, to ask questions and find their own answers, and to develop patience, persistence, and confidence in their ability to tackle and solve real problems.

To produce evaluation criteria for identifying the most effective print instructional materials available, the NSRC drew upon three primary sources:

- the experience of teachers, superintendents, principals, and science curriculum coordinators across the United States;
- the quality standards identified by the NSRC for evaluating units of science instruction in its ongoing review of science curriculum materials under the auspices of the National Academy of Sciences and the Smithsonian Institution; and
- the National Science Education Standards, which were under development at the same time as this resource guide.

The evaluation criteria that NSRC developed were applied in the structured review of curriculum materials. The criteria consist of two sets of questions. The first focuses on pedagogical issues, the second on science issues.

The pedagogical criteria elaborate on the following key questions: (1) Do the materials address the important goals of elementary science teaching and learning? (2) Are inquiry and activity the basis of the learning experiences? (3) Are the topic of the unit and the modes of instruction developmentally appropriate? Additional issues related to presentation and format and to hands-on science materials are then considered.

The set of criteria on science issues expands upon the key questions of whether the science content is accurate, up to date, and effectively presented. It then focuses on aspects of the way science is presented in the materials—for example, whether the writing style is interesting and engaging while respecting scientific language.

The NSRC evaluation criteria are reprinted in appendix B, "NSRC Evaluation Criteria for Curriculum Materials." Teachers, curriculum specialists, curriculum developers, principals, superintendents, and those involved in various aspects of science education reform may find the criteria not only instructive, but useful as an actual review instrument when the need arises to consider the strengths and weaknesses of particular curriculum materials.

The review process developed by the NSRC for the selection of curriculum materials consisted of two phases:

- PHASE I: Teams of experienced teachers and science curriculum specialists reviewed materials for pedagogical appropriateness. Each document received a minimum of two independent reviews. Volumes not recommended in this phase received no further consideration.

Phase I review teams consisted of teachers and science curriculum specialists experienced and knowledgeable in the teaching of elementary school science. Most members of the teams were lead science teachers or master teachers who had taught in school districts with effective science programs. Their backgrounds included participation in numerous science curriculum development activities; they had training and experience teaching children with different learning styles and abilities, and had taught student populations representing diverse cultural and ethnic backgrounds.

Phase I teams included individuals with experience and training in cooperative learning, assessment strategies, the integration of curriculum, and the use of modern technology. Reviewers had experience with a variety of instructional materials for elementary school science programs and were able to use the NSRC evalution criteria effectively to identify differences and to recognize strengths and weaknesses in curriculum materials.

- PHASE II: Scientists reviewed the materials recommended in Phase I to determine if their science content is accurate, current, and presented effectively.

Phase II review teams consisted of scientists with expertise in one of four areas—life science, earth science, physical science, and applied science or technology. Every effort was made to match each scientist reviewer with curriculum materials relevant to his or her area of expertise.

The members of the scientist review teams were teaching professors, working scientists, and others with an understanding of precollege science education. Their involvement with precollege students and science took various forms—for example, in judging science fairs, making classroom presentations about science concepts and careers in science, and sharing their science expertise with classroom teachers. Many of the panel members had experience teaching science at undergraduate and graduate levels; some had taught science courses to future teachers.

Materials that passed review by both the teacher and the scientist review panels are an-

DRAFT

notated in part 2 of the guide. It should be noted that not every individual entry in the guide necessarily meets all the criteria. The NSRC evaluation criteria were designed as a standard to be met, as the ideal level of quality to be sought, and as a working tool that can help inform science curriculum as it is developed. The criteria represent goals—but reachable goals. The curriculum materials included in this guide have accomplished the overall objective of meeting these goals, thereby enhancing the teaching of science through hands-on, inquiry-centered, pedagogically and scientifically sound learning experiences.

The curriculum materials are not ranked or rated here for several reasons. They have all achieved the general objectives set by the criteria. Their inclusion indicates that teachers and scientists have judged them to be effective materials. Beyond that, each item is unique and accomplishes these objectives in its own individual fashion. Ultimately, it is up to teachers and schools to select the particular materials that best fit their needs. Thus, ranking could be misleading—what might be considered exceedingly useful in one classroom might be less so elsewhere because of different needs and circumstances. The full array of materials presented for consideration is meant to offer diversity so that teachers and schools can select what best suits their own needs.

No judgment should be inferred about any elementary science programs, materials, or sources of assistance not included. The guide presents a selected, not an exhaustive, listing of elementary school science curriculum materials.

What Is Not Included in the Guide

Several kinds of teaching resources are not reviewed in *Resources for Teaching Elementary School Science*. Computer software for elementary science, audiovisual materials, science trade books, and elementary science textbooks are not included.

Many excellent science software and audiovisual products exist, can play an important role in the science classroom, and can be integrated with print materials and kits to enrich science teaching. The guide does not undertake to review the vast array of available software programs and audiovisual materials, such as films, videotapes, filmstrips, slides, posters, videodisks, multimedia programs, and so forth. It concentrates instead on print curriculum materials, although some of these also have a software or audiovisual component.

For current information on software and audiovisual products, readers are referred to a software directory and a variety of periodicals and resource guides that feature reviews of audiovisual and computer software materials. (See chapter 7, "Science Book Lists and Resource Guides" and chapter 8, "Periodicals.")

Resources for Teaching Elementary School Science also does not attempt to review the vast number of science trade books available to enrich children's knowledge and understanding. Many teachers use such books as an integral part of their science curriculum, and the NSRC urges teachers to supplement hands-on activities in the classroom with extensive reading. For sources of current information on science trade books, readers are referred to chapters 7 and 8.

And, finally, elementary science textbooks, which typically include few opportunities for meaningful hands-on experiences, are not included. Although textbooks are at times used successfully as supplements to inquiry-based science

DRAFT

programs, the NSRC believes that an elementary science program should not be centered on the use of a textbook alone. Science is a process and a way of thinking. Both aspects require active participation by the individual learner. Students need to be able to carry out scientific investigations using a wide variety of concrete materials, set up their own experiments, change variables systematically, make accurate observations and measurements, and record and graph data.

Getting Started

Readers with differing experience in the teaching of elementary school science will no doubt use this volume. The National Science Resources Center encourages those wanting to get under way with hands-on inquiry-centered science teaching as well as those experienced in this style of teaching to explore the wide array of materials and resources described here.

Research has shown that most children learn science better and sharpen their problem-solving skills most effectively through hands-on instruction. To teachers who are just getting started with this approach, the NSRC recommends that they begin by introducing hands-on units one at a time into their science classes in order to become more comfortable with this style of teaching. Time and again, that experience has encouraged teachers to expand their hands-on teaching, for they see their students learning science in a way that engages them and offers lasting educational benefits.

Children take natural delight in "doing" science. The National Science Resources Center offers *Resources for Teaching Elementary School Science* in the hope that it will encourage more and more teachers to teach hands-on science and that it will help them to do so successfully.

In Press
National Academy Press
1996

DRAFT

APPENDIX B: NSRC EVALUATION CRITERIA FOR CURRICULUM MATERIALS

Consistent with the National Science Resources Center's (NSRC's) philosophy of science teaching and with the recently published *National Science Education Standards* of the National Research Council, the materials included in *Resources for Teaching Elementary School Science* are hands-on and inquiry-centered. Briefly described, such materials provide opportunities for children to learn through direct observation and experimentation; they engage students in experiences not simply to confirm the "right" answer but to investigate the nature of things and to arrive at explanations that are scientifically correct and satisfying to children; and they offer students opportunities to experiment productively, to ask questions and find their own answers, and to develop patience, persistence, and confidence in their ability to tackle and solve real problems.

To produce evaluation criteria for identifying the most effective print instructional materials available, the NSRC drew upon three primary sources:

- the experience of teachers, superintendents, principals, and science curriculum coordinators across the United States;
- the quality standards identified by the NSRC for evaluating units of science instruction in its ongoing review of science curriculum materials under the auspices of the National Academy of Sciences and the Smithsonian Institution; and
- the National Science Education Standards, which were under development at the same time as this resource guide.

The evaluation criteria that NSRC developed were applied in the structured review process of curriculum materials for this guide. These criteria consist of two sets of questions. The first focuses on pedagogical issues, the second on science issues.

The pedagogical criteria elaborate on the following key questions: (1) Do the materials address the important goals of elementary science teaching and learning? (2) Are inquiry and activity the basis of the learning experiences? (3) Are the topic of the unit and the modes of instruction developmentally appropriate? Additional issues related to presentation and format and to hands-on science materials are then considered.

The set of criteria on science issues expands upon the key questions of whether the science content is accurate, up to date, and effectively presented. It then focuses on aspects of the way science is presented in the materials—for example, whether the writing style is interesting and engaging while respecting scientific language.

Two major considerations should be kept in mind when one is using this document:

- The NSRC evaluation criteria provide two gauges for assessing curriculum materials: first, they enunciate specific goals and, second, taken as a whole, they represent the overall level of quality necessary for materials to be effective. Therefore, while materials may not meet each individual criterion completely, they can still reach the overall level of effectiveness defined in the evaluation instrument. That is, if they offer hands-on, inquiry-centered, pedagogically and scientifically sound learning experiences, they may be considered effective even though they do not meet each specific criterion within these categories. The NSRC evaluation criteria were designed as a standard to be met, as the ideal level of quality to be sought, and as a working tool that can help inform science curriculum as it is developed.
- The expectations for core materials are more comprehensive than for supplementary materials. For example, core materials would be expected to provide assessment strategies, whereas science activity books would not. Likewise, core materials would allow students to study a concept in depth, while supplementary materials might provide only a general introduction or isolated activities.

The NSRC evaluation criteria are reprinted in full in this appendix. Teachers, curriculum specialists, curriculum developers, principals, superintendents, and those involved in various aspects of science education reform may find the criteria not only instructive, but useful as an actual review instrument when the need arises to consider the strengths and weaknesses of particular curriculum materials.

DRAFT

NATIONAL SCIENCE RESOURCES CENTER
SMITHSONIAN INSTITUTION • NATIONAL ACADEMY OF SCIENCES
Resources for Teaching Elementary School Science

SCIENCE INSTRUCTIONAL MATERIALS REVIEW FORM

TITLE: *or name of resource*

SERIES TITLE: *if applicable*

AUTHOR(S): *if applicable*

CITY/STATE: *where published*

PUBLISHER/SOURCE:

COPYRIGHT DATE: ISBN NO: ADVERTISED GRADE LEVEL(S): *grade(s)*

SUPPLIES: *availability of materials and kits for core curriculum materials*

COST: *suggested list price*

RESOURCE TYPE: *student activity book, teacher's guide, books on teaching science, etc.*

SUBJECT: *selected from major content categories*

Please supply the following information:

REVIEWER: ______________________ *(reviewer's name)* DATE: ______________ *(date of review)*

RECOMMENDED USER:

(check each that applies) _____ stu _____ tchr _____ adm _____ other (__________)

GRADE LEVEL(S) RECOMMENDED BY REVIEWER IF DIFFERENT FROM THE ADVERTISED LEVEL(S) STATED ABOVE:

(Please circle the specific grade level(s) for which you believe these materials are most appropriate.)

K 1 2 3 4 5 6 7 8 9 10 11 12

Reviewer: ______________________

PEDAGOGY

Instructions: The following questions are designed to help you identify the important elements of each criterion. Please respond by selecting "yes" if the material meets this goal and "no" if it does not. If "no" is selected, please explain the reason in the space provided below the question. In some instances, the question may not be applicable; then mark "NA."

CRITERIA	RATING
I. ADDRESSING THE GOALS OF ELEMENTARY SCIENCE TEACHING AND LEARNING	
Does the material focus on concrete experiences by the children with science phenomena? Reason:	Yes No NA
Does the material enable children to investigate important science concept(s) in depth over an extended period of time (core materials only)? Reason:	Yes No NA
Does the material contribute to the development of scientific reasoning and problem-solving skills? Reason:	Yes No NA
Does the material stimulate student interest and relate to their daily lives? Reason:	Yes No NA
Does the material allow for or encourage the development of scientific attitudes and habits of mind, such as curiosity, respect for evidence, flexibility, and sensitivity to living things? Reason:	Yes No NA
Are assessment strategies aligned with the goals for instruction? Reason:	Yes No NA
Will the suggested assessment strategies provide an effective means of assessing student learning? Reason:	Yes No NA

Reviewer: ______________________

378

DRAFT

NSRC Evaluation Criteria for Curriculum Materials

CRITERIA	RATING
II. FOCUSING ON INQUIRY AND ACTIVITY AS THE BASIS OF LEARNING EXPERIENCES	
Does the material engage students in the processes of science? Reason:	Yes No NA
Does the material provide opportunities for students to make and record their own observations? Reason:	Yes No NA
Does the material provide opportunities for students to gather and defend their own evidence? Reason:	Yes No NA
Does the material provide opportunities for students to express their results in a variety of ways? Reason:	Yes No NA
Does the material provide opportunities for students to work collaboratively with others? Reason:	Yes No NA
Does the material include a balance of student-directed and teacher-facilitated activities? Reason:	Yes No NA

BEST COPY AVAILABLE

Reviewer: ______________________________

CRITERIA	RATING
III. INSTRUCTIONAL APPROACH	
Does the material present a logical sequence of related activities that will help students build conceptual understanding over several lessons? Reason:	Yes No NA
Does the suggested instructional sequence take into account children's prior knowledge and experiences? Reason:	Yes No NA
Are opportunities included to assess children's prior knowledge and experiences? Reason:	Yes No NA
Do the suggested student activities develop critical thinking and problem-solving skills? Reason:	Yes No NA
Does the material incorporate effective strategies for the teacher and/or the students to use in assessing student learning? Reason:	Yes No NA
Does the material incorporate technological applications of science and the interactions among science, technology and society? Reason:	Yes No NA
Do the subject matter and methods of instruction provide suggestions for integrating science with other important learning experiences in the elementary curriculum, such as mathematics, language arts, and social studies? Reason:	Yes No NA

Reviewer: ______________________________

DRAFT

ASSESSMENT OF PEDAGOGICAL APPROPRIATENESS OF MATERIALS

Please provide a *brief* overview of the concepts taught and the activities suggested in this material. It is not necessary to use complete sentences; words and brief phrases are sufficient.

With the above criteria in mind, please comment on any particular strengths in this material.

With the above criteria in mind, please comment on any particular weaknesses in this material.

After reviewing this material with only the above criteria for pedagogical appropriateness in mind, I would:

_______ recommend this material for inclusion

_______ not recommend this material for inclusion

Reviewer: ____

381

PRESENTATION AND FORMAT, MATERIALS, AND EQUITY

Instructions: The following questions are designed to help you identify the important elements of criteria involving presentation and format, materials, and equity issues. Please respond by selecting "yes" if the material meets this goal and "no" if it does not. If "no" is selected, please explain the reason in the space provided below the question. In some instances, the question may not be applicable; then mark "NA."

CRITERIA	RATING
PRESENTATION AND FORMAT	
Teacher materials: Does the background material for the teacher provide sufficient information on the scientific content? Reason:	Yes No NA
Does the background material for the teacher provide sufficient information on common student misconceptions? Reason:	Yes No NA
Is the format easy for a teacher to follow? Reason:	Yes No NA
Are the directions on implementing activities clear? Reason:	Yes No NA
Are the suggestions for instructional delivery adequate? Reason:	Yes No NA
Are the suggested times for instruction reasonable? Reason:	Yes No NA
Student materials: Are the written materials for the students well-written, age-appropriate, and compelling in content? Reason:	Yes No NA

Reviewer: ______________________________

APPENDIX B

DRAFT

NSRC EVALUATION CRITERIA FOR CURRICULUM MATERIALS

CRITERIA	RATING
HANDS-ON SCIENCE MATERIALS	
Teacher materials: Is a master source list of materials provided? Reason:	Yes No NA
Is a list of materials included for each activity? Reason:	Yes No NA
Is a complete set of materials readily available at a reasonable cost? Reason:	Yes No NA
Are refurbishment materials easily obtained and affordable? Reason:	Yes No NA
Student materials: Are the materials recommended for use appropriate for the designated age levels? Reason:	Yes No NA
Are appropriate safety precautions included, where needed? Reason:	Yes No NA
Are instructions on manipulating laboratory equipment and materials clear and adequate? Reason:	Yes No NA
SCIENCE FOR ALL	
Is the material free of cultural, racial, ethnic, gender, and age bias? Reason:	Yes No NA
Are appropriate strategies included/used to meet the needs of special/diverse populations? Reason:	Yes No NA

Reviewer: ______________________________________

383

BEST COPY AVAILABLE

ASSESSMENT OF PRESENTATION AND FORMAT, HANDS-ON SCIENCE MATERIALS, AND EQUITY

With the above criteria in mind, please comment on particular strengths or weaknesses in this material.

After reviewing this material with only the above criteria for presentation and format, hands-on science materials, and equity issues in mind, I would:

_______ recommend this material for inclusion

_______ not recommend this material for inclusion

Reviewer: ________________________________

RECOMMENDATION

Based upon all aspects of my review of this material,

________ I highly recommend this material for inclusion in *Resources for Teaching Elementary School Science.*

________ I recommend this material for inclusion in *Resources for Teaching Elementary School Science.*

________ I recommend this material for inclusion in *Resources for Teaching Elementary School Science* with reservations.

Primary reason for reservations: ________________________________

____

____

________ I do not recommend this material for inclusion in *Resources for Teaching Elementary School Science.*

Primary reason for rejection: ________________________________

____

____

Reviewer: ________________________________

NATIONAL SCIENCE RESOURCES CENTER
SMITHSONIAN INSTITUTION • NATIONAL ACADEMY OF SCIENCES
Resources for Teaching Elementary School Science

SCIENCE INSTRUCTIONAL MATERIALS REVIEW FORM

TITLE: *or name of resource*

SERIES TITLE: *if applicable*

AUTHOR(S): *if applicable*

CITY/STATE: *where published*

PUBLISHER/SOURCE:

COPYRIGHT DATE: ISBN NO: ADVERTISED GRADE LEVEL(S): *grade(s)*

SUPPLIES: *availability of materials and kits for core curriculum materials*

COST: *suggested list price*

RESOURCE TYPE: *student activity book, teacher's guide, books on teaching science, etc.*

SUBJECT: *selected from major content categories*

The material you are reviewing has already been identified by teachers and science curriculum specialists, in a comprehensive review process, to be pedagogically effective instructional material that would support a "hands-on, constructivist, inquiry-based" elementary-school science program. Your task is to review the material (including the background information for teachers) to evaluate the science content for its accuracy and currency, and the effectiveness of its presentation.

Please supply the following information:

REVIEWER: ______________________________ DATE: ______________________
(reviewer's name) *(date of review)*

Reviewer: ______________________________

SCIENCE CONTENT, PRESENTATION, AND EQUITY

Instructions: The following questions are designed to help you identify the important elements of each criterion. Please respond by selecting "yes" if the material meets this goal and "no" if it does not. If "no" is selected, please explain the reason in the space provided below the question. In some instances, the question may not be applicable; then mark "NA."

CRITERIA	RATING
SCIENCE CONTENT	
Is the science content incorporated in the materials accurately represented? Reason:	Yes No NA
Is the science content consistent with current scientific knowledge? Reason:	Yes No NA
Are important ideas included? Reason:	Yes No NA
Are generalizations adequately supported by facts? Reason:	Yes No NA
Are facts clearly distinguished from theories? Reason:	Yes No NA
Do the suggested investigations lead to an understanding of basic principles? Reason:	Yes No NA
Do experiments and activities promote student understanding of how scientists come to know what they know and how scientists test and revise their thinking? Reason:	Yes No NA

Reviewer: ______________________________

387

CRITERIA	RATING
SCIENCE PRESENTATION	
Is science shown to be open to inquiry and controversy and free of dogmatism? Reason:	Yes No NA
Are different scientific viewpoints presented when appropriate? Reason:	Yes No NA
Are personal biases avoided? Reason:	Yes No NA
Is the writing style interesting and engaging, while respecting scientific language? Reason:	Yes No NA
Is vocabulary used to facilitate understanding rather than as an end in itself? Reason:	Yes No NA
Is science represented as an enterprise connected to society? Reason:	Yes No NA

CRITERIA	RATING
SCIENCE FOR ALL	
Is material free of cultural, racial, ethnic, gender, and age bias? Reason:	Yes No NA

Reviewer: ______________________________

DRAFT

ASSESSMENT OF SCIENCE CONTENT, PRESENTATION, AND EQUITY

With the above criteria in mind, please comment on any particular strengths in this material.

With the above criteria in mind, please comment on any particular weaknesses in this material.

RECOMMENDATION

After reviewing this material with the above criteria for science content and presentation in mind, I would:

_______ highly recommend this material for inclusion in *Resources for Teaching Elementary School Science.*

_______ recommend this material for inclusion in *Resources for Teaching Elementary School Science.*

_______ recommend this material for inclusion in *Resources for Teaching Elementary School Science* with reservations.

Primary reason for reservations: ____________________

_______ not recommend this material for inclusion in *Resources for Teaching Elementary School Science.*

Primary reason for rejection: ____________________

Reviewer: ____________________

MEMORANDUM

FROM: John Monk and Todd Fennimore
TO: SchoolNet Staff
RE: Progress report on SchoolNet Software Review Project
DATE: October 2, 1996

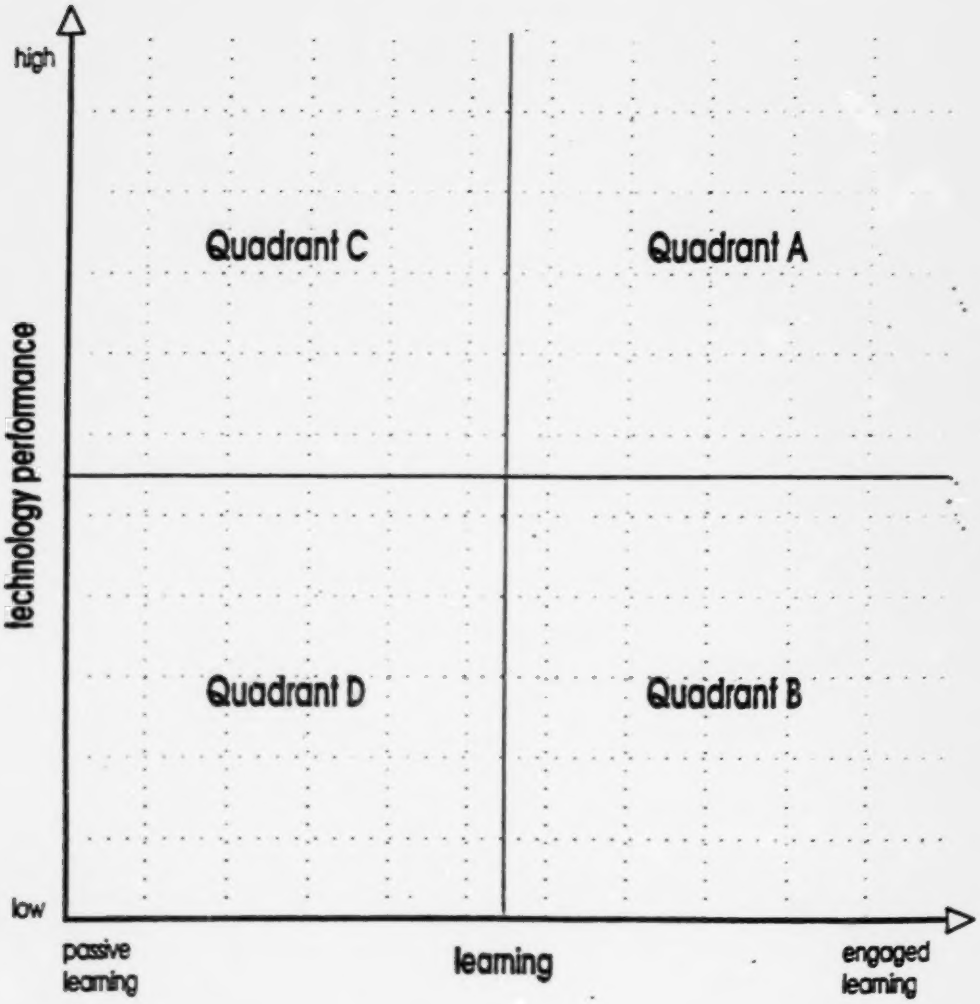
PROGRESS REPORT ON SCHOOLNET SOFTWARE REVIEW PROJECT (SSRP)

Background to the SchoolNet Software Review Project (SSRP)

In 1995, the North Central Regional Educational Laboratory (NCREL) and the Council for Educational Development and Research (CEDaR) published a document called *Plugging In*, authored by Jones, Valdez, Nowakowski, and Rasmussen. This document helped educators assess, in a systematic way, the extent to which technology is currently used in their schools to engage active and meaningful learning, and plan for changes in their use of technology that would move them further in this direction. A critical insight fostered by the use of *Plugging In* is the following: It is possible to make advanced use of technology to support passive, largely teacher-directed learning, much as it is possible to lack technological power but, nonetheless, engage students in active, student-directed learning. This insight was crystallized in a graphic that represented visually the results of using the *Plugging In* instrument; four quadrants represented the range of possibilities: placement in Quadrant A indicates engaged learning and high technology performance; Quadrant B, engaged learning and low technology performance; Quadrant C, passive learning and high technology performance; and Quadrant D, passive learning and low technology performance [see graphic on next page]. This visualization has proven powerful and useful to educators since, at a glance, they can see where they are and where they want to go as they employ the tools of technology to invigorate learning.

The *Plugging In* instrument, then, was designed to be an instrument to help educators plan their trajectory of change in use of technology to promote engaged learning and to develop a richer understanding of the ways that technology in general can be used to engage active and meaningful learning. As educators began using this instrument in the field, further needs were identified and targeted for what can be seen as the subsequent generations of the *Plugging In* instrument. The first need identified was the desire for an instrument that would speak to the contextual contingencies of the classroom and content-specific learning goals, standards, and approaches. The second need was for instruments (surveys, questionnaires, matrices, etc.) that would address particular types of technology, such as authoring/multimedia tools, navigation tools for

Table 3: The Learning and Technology Framework


high
technology performance
low
Quadrant C
Quadrant A
Quadrant D
Quadrant B
passive learning
learning
engaged learning

electronic reference and information, instructional software, learning management tools, and networking and telecommunications technologies. There was also a general need for linking the results of these evaluations to school-based and administrative decisionmaking and strategizing.

Awareness of these needs arose, in part, from work with Jamison CompuTech, an elementary school of the Cleveland Public School District, and the Reynoldsburg School District, especially with faculty from Reynoldsburg High School and Taylor Road Elementary School, as they learned about and used *Plugging In*. WVIZ, the PBS station from Cleveland, in collaboration with NCREL, recently completed a broadcast-quality videotape that captures the work of these schools as they use *Plugging In* to accomplish their own authentic tasks. In the case of Cleveland, a broad-based team from the school employed the instrument to help them assess needs in technology and plan for changes to meet them. In the case of Reynoldsburg, select faculty from both the high school and the elementary school were invited to join Todd Fennimore in a collaboration to adapt the *Plugging In* instrument for use in generating evaluations of instructional software for students in terms of its responsiveness to content standards, the contextual realities of the classroom, and potential to engage students in authentic tasks that build conceptual understanding. This task was carried out over late Spring and early Summer, in two afternoon inservices and one planning meeting during the school year and in one day-long pre-planning meeting and three full-day working sessions in July. The result was a respectable working draft of a software evaluation protocol developed *by* teachers, *for* teachers. After further external reviews were made, and recommended changes implemented, a version suitable for use at a summer institute devoted to software evaluation was completed.

Need for the SchoolNet Software Review Project (SSRP) and Plan of Work

This effort to develop, and subsequently disseminate, a software evaluation protocol meshes well with the needs of SchoolNet Plus. The overarching goal of SchoolNet Plus is to ensure that there is at least one computer for every 5 students in K-4 classrooms and that these resources are used to engage active and challenging learning. SchoolNet staff felt that, if educators were to use this new capacity to promote engaged learning for all students, they must develop a database of information on software which helps teachers accomplish this, as well as the tools for evaluating software themselves in terms of its capacity to engage learners in authentic tasks and conceptually meaningful, challenging problems. Ohio SchoolNet has given Eisenhower National Clearinghouse on Mathematics and Science Education this grant to develop a protocol that educators could use to evaluate software in terms of its ability to immerse students in a rich, problem-solving environment; convene a group of K-4 educators in the first year of the two-year grant to evaluate software for K-4 mathematics and science education; support this group of teachers trained in the use of the protocol to organize and conduct professional development experiences in their own region of Ohio. A similar cycle of activities will be conducted to evaluate K-4 software in social studies and language arts, create a database of these evaluations, and help other educators in Ohio become aware of the initiative and use the protocol developed in making decision about their own software purchases. The rest of this report indicates the progress that has been made on this plan of work.

Chronology of Activities from May 1996 to September 1996

Writing and gaining approval for grant. A preliminary proposal was submitted by Dr. John Monk for Dr. Timothy Best's review on March 13, 1996. Based on reviews of the draft by Dr. Best and Dr. Beau Fly Jones, a revised and final proposal was submitted by Dr. Monk on April 24, 1996. Approval for the project and grant moneys were received in May of 1996.

Conceptual groundwork laid in work with Reynoldsburg Public Schools. Todd Fennimore conducted two afternoon inservices with faculty from the Reynoldsburg Public Schools in early May to familiarize them with *Plugging In* and its use in assessing the desirability of technologies in terms of engaging learning. Another day-long meeting on May 23, 1994 was conducted to set out parameters of the task of adapting *Plugging In* to the purpose of evaluating software. (A list of those involved from Reynoldsburg Public School is provided on the first page of the software evaluation protocol, which is attached.)

Call for applicants to the Summer Institute through the Regional Communities of Practice and the selection process. A description of the SSRP project and an initial call for applicants was made to directors of the Regional Communities of Practice at a May meeting convened by Jenny Moormeier of Ohio SchoolNet. This was followed up a week later by a faxed call for applicants (see attached document), which detailed the criteria for selection of candidates, specified materials to be included in an application, and set a deadline of June 7, 1996 for receipt of applications. (This deadline was later extended to June 12, 1996).

Six criteria were used to select candidates: 1) extensive experience in teaching or working with teachers in a K-4 setting; 2) demonstrated working understanding of hands-on, conceptually oriented approaches to learning in general; 3) strong knowledge base regarding both the content and pedagogy of mathematics and science; 4) adequate comfort level with using computers; 5) preparedness in assuming leadership in organizing and conducting professional development activities in their region; and 6) broad regional representation. Over 25 strong applications were received. Those sixteen applicants who provided concrete and vivid demonstrations of using constructivist approaches in the classroom and ensured broad regional representation were selected. Judy Wahrmann (then co-director of the Northeast RCOP) and Zana Vincent (co-director of the Central Ohio RCOP) provided feedback on candidates selected and their regional representation. (A list of participants and their schools is attached.)

Refinement of protocol done over the summer by faculty of Reynoldsburg Public Schools along with critical friend reviews. Dr. Beau Jones, Dr. Colleen Sexton, Steve Dackin (principal of Reynoldsburg High School), Sue Dackin (elementary teacher at Taylor Street Elementary School in Reynoldsburg), and Marcy Raymond (chemistry teacher at Reynoldsburg High School) met to lay the groundwork for the format of the instrument and to brainstorm content that fit into that format. This, in addition to typing and reviewing the notes generated in the later school year meeting, became grist for the three days of working meetings on July 9th, 10th, and 15th of 1996, the product of which was a working draft of the software evaluation protocol. This protocol was reviewed and refined by Dr. Jones, Dr. Monk, and Todd Fennimore. Further external reviews were provided by Dr. Sexton and Dr. Fred Whiteman. These efforts resulted in a good working version of the protocol for use at the Summer Institute on Software Evaluation for K-4 Mathematics and Science Educators. Testimony that a teacher-friendly instrument was developed is given by the fact that only minor revisions to the instrument were suggested (and implemented) by the Summer Institute participants.

Planning and implementing the Summer Institute. As a first step in planning the Institute, all teachers were contacted individually to confirm their commitment to attending the Institute during the week of June 17-21, 1996. This phone communication was followed up by a letter shortly thereafter confirming their selection, informing them of logistical and payment details, and giving them more background information on the SSRP project and the Eisenhower National Clearinghouse. They then received a more substantial information packet in July that included more logistical details (e.g., hotel arrangements, directions to ENC), a draft of the software evaluation protocol, and copies of relevant parts of the national and state standards in mathematics and science to better prepare them for the Institute. John Monk and Todd Fennimore met with ENC staff to apprise them of the needs of the Summer Institute participants for assistance in applying standards and computer help. Apple and DOS-based machines were rented and set up for the Institute.

The first two days of the Institute (August 5-6) were spent familiarizing teachers with the national and state standards that they were to apply in evaluating software titles, as well as working through the application of the protocol to a software title (*Sammy's Science House*) to discuss the way to interpret various items on the protocol, surface any difficulties, and come to working consensus on use and understanding of the protocol. The other days (August 7-9 and August 12-16) were devoted solely to evaluating software titles in K-4 mathematics and science. On the last day of the Institute, teachers finished their reviews in the morning, then debriefed on their experiences in using the protocol, recommended final improvements to it, and strategized ways to mount professional development initiatives in their own and adjacent regions. During the course of the Institute, over 200 software titles were evaluated. These evaluations are in electronic form and will be available on-line in early October.

Initial efforts to assist Summer Institute participants in delivering professional development in their own and adjacent regions. Meetings have already been scheduled with a number of Summer Institute participants and those who will be their partners in providing regional professional development experiences. Todd Fennimore will be discussing next steps for the Regional Communities of Practice to take with trained teachers on September 29, 1996. On October 3rd, 1996, a meeting organized by Fred Whiteman will take place, bringing together key technology and learning players in the Central Ohio region, including representatives from CORPDC and ITSCO, teachers who participated in the Summer Institute, and ENC staff to plan professional development initiatives linking learning and technology. Other meetings planned are with Kay Deitchel, the curriculum supervisor for the Scioto County Educational Service Center and a Summer Institute participant, who will be working on software adoption plans with administrators and select teachers in her region and with Carol Collins, science education specialist at Hamilton County Educational Service Center, to look at ways to tap the expertise of the SSRP-trained teachers as she plans workshops in software selection and evaluation. Phone follow-up with all Summer Institute participants will be conducted over the course of October, 1996.

Summary of Achievements to Date

Development of *Learning through Software* evaluation protocol. The respectable product generated from co-development with teachers is a powerful illustration of what can come of a genuine, ongoing collaboration between practitioners and researchers. The final version is attached.

Successful Summer Institute. The Directors of the Regional Communities of Practice were highly effective in identifying highly qualified and energetic teachers throughout Ohio to participate in the Summer Institute. The Summer Institute participants performed their evaluations with enthusiasm and intelligence. They had a positive assessment of the Summer Institute, and feel well-positioned to provide professional development experiences, working in partnership with the Regional Communities of Practice, in their own and adjacent regions. (See attached summary of teacher evaluations of the Summer Institute.)

Attachment of evaluation summaries to ENC catalog records for software. During the first 2 weeks of October, a summary for each software item reviewed during the Summer Institute will be placed in the ENC catalog. Items containing these summaries may be located using the ENC resource finder searching under words for the acronym SSRP, and will be searchable using a special resource finder which will only search those items reviewed by the institute participants.

Database of teacher-generated software evaluations in electronic form. The software evaluations are in electronic form, and will be made available online either through the ENC web site, or through the web site being developed for SchoolNet. We anticipate this being completed sometime in October.

FileMaker database of reviews available. For individuals associated with the project who have access to FileMaker 3.0, a special database containing the reviews generated by the summer institute will be made available in mid October.

Special ENC Focus issues to be developed. A special issue of ENC focus will be developed during October for release in early November (depending upon time to print issues). This issue will highlight the highest rated software reviewed by the Institute participants. Plans are to deliver 30,000 to 50,000 copies of this document, depending upon current printing costs, to SchoolNet staff for subsequent distribution throughout the state.

Language arts and social science activities. Project staff are now beginning to gather social studies and language arts software. The search for the cataloger to work with this software will be initiated during October, as will efforts to identify the language arts and social studies resource specialists who will assist in the development of instrumentation and the expansion of the ENC collection in these areas. Staffing for this portion of the project is anticipated to be completed some time in late October or early November.

ATTACHMENTS: Version 2.1 of Protocol, call for applicants, list of Summer Institute participants and their schools, summary of Summer Institute evaluation results

Ohio SchoolNet Plus

Learning Through Software

A Software Evaluation Protocol Designed For Teachers By Teachers

Version 2 – August 28, 1996
Draft – Not for general release – do not duplicate

This protocol was developed by teachers of the Reynoldsburg City School District, Reynoldsburg, Ohio.

Reynoldsburg High School	Taylor Road Elementary School
David Baker, Mathematics Teacher	Susan Dackin, 1st grade Teacher
Charles Brads, Mathematics Teacher	Marlie Griffin, Special Education Teacher
Steve Dackin, Principal	Wayne Kanzigg, 3rd grade Teacher
Richard Ladowitz, Biology Teacher	Dawn Tufto, 4th grade Teacher
Marcy Raymond, Chemistry Teacher	

Additional assistance was provided by:
Deborah Bergeron, Principal, Taylor Road Elementary School
William Gathergood, English Teacher, Reynoldsburg High School
Lanna Trimmer, Kindergarten Teacher, Taylor Road Elementary School

This protocol was developed as part of the SchoolNet Plus Software Review Project (ENC/SSRP) facilitated by the Eisenhower National Clearinghouse for Mathematics and Science Education, and funded by the SchoolNet Plus Project, Ohio Department of Education. Project Coordinator: Todd Fennimore, Senior Research Associate, Ohio Supercomputer Center.

To make best use of this protocol, it is important to be familiar with and make reference to the appropriate national and state standards and guidelines.

Mathematics:	Curriculum and Evaluation Standards (1989) National Council of Teachers of Mathematics
	Model Competency-Based Mathematics Program (1990?) Ohio Department of Education
Science:	National Science Education Standards (1995) National Research Council
	Science: Ohio's Model Competency-Based Program (1994) Ohio Department of Education

Part I: Software Profile

Title of Software ______________________________

Version ________________

Publisher ______________________________

ENC-Number ________________

Primary Subject Focus (circle one):
Mathematics
Science
Interdisciplinary/Integrated Math and Science

Additional Subjects Addressed (circle all that apply):
Mathematics
Science
Language Arts
Social Studies
Other: ____________

This software can be used for (check all that apply):

___ Remediation/Review
___ Tutorial
___ Assessment
___ Simulation
___ Demonstration
___ Authoring (e.g., drawing, publishing, multimedia production)
___ Other ____________________
___ Drill and practice
___ Enrichment
___ Game
___ Collaborative projects
___ Microcomputer-based laboratory
___ Standard Instruction
___ Information retrieval/resource
___ Instructional Management
___ Problem solving

This package is designed to be used with (circle all that may apply):

Individual students   Small Groups (2-15)   Whole classes (16 or more)

The predicted level of interest of students in this software is:

Low   Moderate   High

This software assessment tools of some kind.

Yes   No

This package incorporates special features which support adaptive/assistive technologies (speech synthesizing, sip-and-puff switch, etc.)?

Yes   No   Insufficient Information

This package is known to work with adaptive/assistive technologies run in conjunction with but not part of the package.

Yes   No   Insufficient Information

Is this package free of stereotypes? Consider the following:

Are females represented in lead and professional roles as frequently as males?

Yes   No   Not applicable

Are non-whites portrayed in lead and professional roles frequently?

Yes   No   Not applicable

Is a diversity of cultural traditions and practices represented in a positive fashion?

Yes   No   Not applicable

Are individuals with physical impairments represented in a positive fashion?

Yes   No   Not applicable

Part II: Evaluator Profile

This evaluation was prepared by (Check one):

_____ An individual teacher
_____ A group of teachers

The individual or group providing this evaluation has (Check one):

_____ No experience using this package with students
_____ Limited experience using this package with students
_____ Extensive experience using this package with students

The grade levels of the teacher (or teachers) doing this evaluation were (check all that apply):

Pre K K 1 2 3 4 5 6 7 8 9 10 11 12

Date of Evaluation: ____/____/____

Optional: If you have used this package with students and would be willing to correspond about or discuss this package with other teachers, please provide the following information:

Name: ______________________________

Address: (If you are willing to correspond about this item)

____

____

Phone: (If you are willing to be contacted by phone)

Home: (_____) _____-________

School: (_____) _____-________

E-Mail: (If you are willing to be contacted by e-mail)

E-Mail: ______________________

Part III: Software Evaluation

A. What do students learn through the software?

1. Is the approach to content taken by the package aligned with standards?

Use the space below to list key standards with which the software aligns. Refer to the Curriculum and Evaluation Standards of the National Council of Teachers of Mathematics, the National Science Education Standards from the National Research Council, and the Ohio Competency-Based Curriculum Models for Mathematics and Science, K-4. Also, rank the level to which the package appears to align with these standards identified using the following scale

Alignment
1 – Poorly aligned
2 –
3 – Moderately aligned
4 –
5 – Well aligned

Source: NCTM, NRC, Ohio	Standard Number	Standard Description or Text	Level of Alignment

Use additional space on the back of this form if necessary

2. To what extent does the software present accurate and up-to-date information?

1	2	3	4	5
Information is frequently inaccurate or out-of-date		Information is occasionally inaccurate and out-of-date		Information is generally accurate and current

3. To what extent does the software allow students to choose among challenging options with multiple solution paths?

1	2	3	4	5
Only simple tasks with one solution option are provided		Some problems are provided that offer more than one solution path		Many challenging problems with multiple solution paths are provided as options to students

4. To what extent does the software make connections among concepts and/or across disciplines?

1	2	3	4	5
Focus is on discrete concepts		In some instances multiple concepts and multidisciplinary perspectives are provided		Strong connections are made among multiple concepts and across multiple disciplines

5. To what extent does the software use problem-solving processes to help students build their conceptual understanding?

1	2	3	4	5
Focus is on drill and practice and/or skill development		Well-defined problems focus on mid-level concepts		Open-ended problem solving processes are used to explore higher level concepts

6. To what extent does the software present concepts in context and in an a clear and understandable manner?

1	2	3	4	5
Concepts are unclear and presented out of context		Concepts presented in a accurate, factual manner with little or no context		Concepts are clear and well-defined and presented within a rich context

7. To what extent does the package present concepts in a manner that facilitates their transfer to other contexts such as written work, classroom activities, and projects?

1	2	3	4	5
Concepts do not transfer		Concepts transfer to some different contexts		Concepts transfer to many different settings

B. How do students learn with the software?

1. To what extent does the software present authentic problems to be solved?

1	2	3	4	5
No problems are presented or problems are contrived and do not reflect real-life situations		Problems are based on real-life situations but are constrained due to the approaches taken by the software		Problems are based on real-life situations and evoke varied and challenging approaches to solution

2. To what extent does the software allow for varied approaches to learning (linguistic, logical-mathematical, spatial, musical, etc.)?

1	2	3	4	5
Software follows one approach to learning		Software supports some varied approaches to learning		Software supports many approaches to learning

3. To what extent does the software motivate and enable the student to construct his or her own model or simulation in the course of problem solving?

1	2	3	4	5
The student is not given the opportunity to develop models or simulations		The student is allowed some flexibility in developing or modifying limited models or simulations		The student is given tools and support for designing his or her own models or simulations

4. To what extent does the software engage students in inquiry around open-ended problems?

1	2	3	4	5
No opportunities for open-ended inquiry are provided		Engagement occurs through solving well defined, stepwise problems		Engagement largely occurs through inquiry focusing on concepts

BEST COPY AVAILABLE

5. To what extent does the software enable students to link concepts in a meaningful sequence while conducting inquiry?

1	2	3	4	5
Primary focus is on skills and facts with little or no meaningful sequence		Concepts are approached as a structured sequence of skills and facts with minimal inquiry		Concepts explored through an open-ended inquiry process that allows the discovery of links among skills and facts

6. To what extent does the software encourage and support collaborative learning?

1	2	3	4	5
Software designed to be used individually		Software allows for cooperation among students on well structured tasks		Software supports collaboration among students in open-ended problem solving settings

7. To what extent does the software allow for the student to review a previous section of the program as needed?

1	2	3	4	5
Software does not allow student to review and/or redo prior work		Software allows students only limited review and/or revision of prior work		Software allows extensive review and/or revision of prior work as needed

8. To what extent does the software encourage the student to assume the role of a self-directed learner?

1	2	3	4	5
Software promotes only teacher-directed and student receptive learning		Software employs some student-directed learning with the teacher remaining in the role of director and authority		Software promotes student-directed learning while encouraging teachers to assume the role of a co-investigator, facilitator, and mentor

C. What are the technical characteristics of the software

1. To what extent does the system provide security for the sections of the software designed for teacher use, which contain personal information about students, or which contain assessment information?

1	2	3	4	5
No security is provided for teacher-only sections of the software or security is easily breached		Modest security is provided but the security could be overcome by a determined student		Security systems provided are highly effective and cannot be easily overcome by students

2. To what extent does the system respond well to student errors or intentional attempts to disrupt software operation?

1	2	3	4	5
Software can easily be disrupted by and recovers poorly from student errors or intentional attempts to disrupt operation		Software can be disrupted with some effort, but in general is tolerant of student errors and intentional attempts to disrupt operation		Software cannot be disrupted by student errors or attempts to disrupt software operation

3. To what extent can students learn to use the software on their own?

1	2	3	4	5
Students cannot use the program without direct assistance from the teacher		Students can use the software relying primarily on included help and documentation with some assistance from the teacher		Students can use the program independently making use of provided help and documentation

4. To what extent does the software provide for intelligent feedback adapted to the student needs?

1	2	3	4	5
Little or no feedback is provided		Feedback provides limited explanations of possible sources of error when errors occur		Feedback analyzes student misconceptions and difficulties when problem solving is unsuccessful

5. To what extent does the software effectively use multimedia (sound, graphics, video, etc.)?

1	2	3	4	5
Multimedia not employed or not relevant to the task undertaken		Multimedia is employed and of interest, but is not critical to the learning of content		Multimedia approaches are appropriately employed to enhance the learning experience

BEST COPY AVAILABLE

6. To what extent does the software provide tools integrated into the program which allow students to collect, analyze, and manipulate data?

1	2	3	4	5
No tools are provided or the tools which are provided do not function well.		Limited tools are provided and generally function in an appropriate fashion.		Extensive tools are fully accessible at all times in the program and their use is encouraged and supported

7. To what extent does the software have a high level of interactivity?

1	2	3	4	5
Software requires little or no interaction by students		Software requires occasional input by student and generally requires only simple responses		Software requires substantial and frequent input from students

8. To what extent does the software integrate assessment throughout the program?

1	2	3	4	5
Assessment is not present or is poorly integrated		Assessment appears only at defined points in the instruction		Assessment is ongoing and integrated with instruction

9. To what extent does the software allow the teacher to modify the software parameters (set difficulty levels, input word lists, etc.) for use with diverse students?

1	2	3	4	5
Teacher cannot modify software parameters, or modification is too difficult		Only limited teacher modifications are available, or modification requires some effort		Teacher can easily and extensively modify the software parameters

10. To what extent does the grade level reported by the software publisher accurately reflect the grade level at which the software should be used?

1	2	3	4	5
The grade levels reported by the publisher appear to be inaccurate		The grade levels reported are appropriate		The grade levels reported are appropriate, but the software could be used for a broader range of students

At what grade levels would it be most appropriate to use this software? (circle all that apply):

Pre K K 1 2 3 4 5 6 7 8 9 10 11 12

11. To what extent does the software contain informative and useful help features?

1	2	3	4	5
Little or no help is provided, or the help which is provided is of little use		Adequate help is provided for the general user		Extensive help, including such things as examples, templates, wizards to guide users, and cross referencing and topic linking are provided to assist users in specific tasks

12. To what extent does the software have an easy installation process?

1	2	3	4	5
Installation was difficult and confusing, and or was unsuccessful on one or more attempts		Installation was moderately easy with adequate instructions provided		Little effort was needed to install the package and clear instructions were provided.

13. To what extent does this package provide effective teaching and learning guides to help teachers integrate the software into the curriculum?

1	2	3	4	5
No guides are provided, or guides provided are ineffective		Moderately effective guides are provided		Highly effective guides are provided

BEST COPY AVAILABLE

D. What are the characteristics of the types of assessments provided by the software?

Note – leave these items blank if no assessment is provided

1. To what extent does the assessments offered identify areas where further student development is needed and suggest paths for further development?

1	2	3	4	5
Software provides no assessment to aid student development		Software identifies areas for further development but offers no means for obtaining that development		Software identifies areas where further development is needed and provides activities or other means for obtaining that development

2. To what extent does the assessment correlate with the learning objectives of the software?

1	2	3	4	5
There is little or no correlation between assessment and learning objectives		The assessment provided represents some of the learning objectives		The assessment represents most or all of the learning objectives

3. To what extent does the software support performance-based assessment by allowing the student to demonstrate his or her knowledge using tools to gather, present, and interpret data?

1	2	3	4	5
Software provides no such tools		Software provides tools which can only be used with data generated by the software		Software provides tools which can be used with data both generated by the software and independently by the student

4. To what extent does the software adapt itself to the skill level of the student?

1	2	3	4	5
Only a set path is provided based on student skills		Some flexibility in movement is based on student skills		Many pathways are provided based on student skills

5. To what extent does the software help students and teachers monitor and track student learning?

1	2	3	4	5
No monitoring or record-keeping provided		Records are kept on student progress for teacher use only		Students and teachers both can track and monitor student learning

6. To what extent does the software accurately begin the student at his or her current level of understanding and skill and allow flexibility for the student?

1	2	3	4	5
The same starting point is provided for all students		Student is placed at an appropriate level by software without the ability to select different starting points		Software starts student at an appropriate level given their understanding and skill with student having the option of adjusting that level

E. Summary Comments

Please provide a brief statement of the strengths and weaknesses you see in this software product. Focus your comments on things that you feel would assist another teacher in determining if this package is right for use in their class with their students. Also, if you have used this software with students, please share your experiences here. Use the back of this page as necessary.

SchoolNet Software Review Project
Summer Institute Evaluation

Summary Data

Please fill this form out to the best of your ability: Feel free to comment on any question if appropriate, and make any additional suggestions or comments at the end of the evaluation form.

What you can do as a result of this Institute:

Scale:	5	I am confident I can do this with a high level of expertise
	3	I am confident I can do this, I just am not sure how well
	1	I doubt that I will be able to do this

Identify educational software characteristics that indicate alignment with national standards in mathematics and science.

1 2 3 4 5 Comments: Mean: 4.588 SD: .618

I have really grown in this expertise. For this and the next two questions: It would have been very helpful to discuss/review the 4 pieces of software we all reviewed to see how I compared with others.

Identify educational software characteristics that indicate software alignment with the Ohio model curricula in mathematics and science.

1 2 3 4 5 Comments: Mean: 4.765 SD: .437

Identify educational software characteristics that indicate software can be used by students to learn mathematics and science in preparation for the Ohio competency tests in mathematics and science.

1 2 3 4 5 Comments: Mean: 4.706 SD: 0.470

Identify the content oriented features that software might possess and how these features reflect national and state standards and guidelines.

1 2 3 4 5 Comments: Mean: 4.588 SD: 0.507

Identify the features of software which support active and meaningful learning.

1 2 3 4 5 Comments: Mean: 4.824 SD: 0.393

Assess the technical features of educational software to determine if it is appropriate for K-4 mathematics and science instruction.

1 2 3 4 5 Comments: Mean: 4.647 SD: .493

Use the *Learning Through Software* evaluation instrument as a means to organize my review of educational software

1 2 3 4 5 Comments: Mean: 4.824 SD: 0.393

Teach others how to interpret the results obtained from the *Learning Through Software* evaluation protocol.

1 2 3 4 5 Comments: Mean: 4.412 SD: 0.618

Teach others how to use ENC to find educational software evaluated by the SSRP project.

1 2 3 4 5 Comments: Mean: 4.412 SD: 0.618

About the Institute in General:

Scale	5	Strongly Agree
	3	Undecided
	1	Strongly Disagree

The institute was well managed.

1 2 3 4 5 Comments: Mean: 5.00 SD: 0

To the extent possible due that many changes needed to be made.

The sessions dedicated to learning about the *Learning Through Software* protocol clearly presented the instrument.

1 2 3 4 5 Comments: Mean: 4.765 SD: 0.437

Not if you were already familiar with them.

The sessions on State and national standards in science were presented in a clear and useful manner.

1 2 3 4 5 Comments: Mean: 4.471 SD: 0.624

The sessions on State and national standards in mathematics were presented in a clear and useful manner.

1 2 3 4 5 Comments: Mean: 4.438 SD: 0.629

The time spent doing a sample software evaluation using the *Learning Through Software* protocol as a group was well spent.

1 2 3 4 5 Comments: Mean: 4.882 SD: 0.332

This was necessary.

The session on using ENC online to find materials was useful.

1 2 3 4 5 Comments: Mean: 3.882 SD: 1.054

Would have liked to actually try on my own. Some tasks to develop skills would have been helpful.

Throughout the institute I felt that I could contribute to the development of the *Learning Through Software* protocol, as well as just doing software evaluations.

1 2 3 4 5 Comments: Mean: 4.765 SD: 0.562

We were made to feel that our comments were welcome.

BEST COPY AVAILABLE

About The Software Review:

Scale 5 Strongly Agree
3 Undecided
1 Strongly Disagree

The pace my partner and I maintained while evaluating software seemed to be appropriate.

1 2 3 4 5 Comments: Mean: 4.765 SD: 0.437

For us it was.

My partner and I contributed equally to the evaluation of the software we reviewed.

1 2 3 4 5 Comments: Mean: 4.824 SD: 0393

We each settled into our own strengths and complemented each other.

The support provided by ENC staff to assist me in loading and using software was appropriate.

1 2 3 4 5 Comments: Mean: 4.941 SD: 0.243

Very helpful and nice. Very supportive and friendly. rob and Ron were excellent role models for how we should help peers back in the district. They were always calm, understanding, and helpful.

The setting provided at ENC for the review of software was conducive to the generation of quality reviews.

1 2 3 4 5 Comments: Mean: 4.941 SD: 0.243

Our needs were met nicely.

The selection of software provided for review was representative of what I know to be available.

1 2 3 4 5 Comments: Mean: 4.529 SD: 0.624

About Institute Logistics:

Scale 5 Strongly Agree
3 Undecided
1 Strongly Disagree

The hotel rooms and service arranged for me (if I stayed at the hotel) were of reasonable quality.

1 2 3 4 5 Comments: Mean: 4.077 SD: 1.188

Hotel rooms were the pits; service rated 5.

Morning coffee and pastries, lunches, and breaks provided at ENC were of reasonable quality.

1 2 3 4 5 Comments: Mean: 4.750 SD: 0.577

Paperwork associated with participating in this institute seemed appropriate.

1 2 3 4 5 Comments: Mean: 4.750 SD: 0.477

BEST COPY AVAILABLE

ENC staff were positive about the work we were undertaking and facilitative of that work.

1 2 3 4 5 Comments: Mean: 5.00 SD: 0

This must be a great place to work for! Everyone made us welcomed. As I walked through the building, the personal relations between employees seemed to be exceptionally pleasant. I found everyone to be very helpful and understanding whether the problem was minor or a major one. Thank you for a very pleasant and educational experience.

Additional Comments:

Thanks to the entire ENC staff.
Your hospitality and planning was superb. Thank you.
I learned a lot. Thank you for this opportunity.
Everyone was nice and helpful.
Well organized. Good help provided. Safe environment.
I learned a lot. Thank you for this opportunity.
Everyone was nice and helpful.
I have grown professionally and as an individual. Thanks for this great opportunity.
I look forward to hearing about the language arts & social studies institute next year.
This has been a great opportunity!
Working through the older software was frustrating at times, but I feel that was a valuable to the project because we were able to document the problems.
There needed to be more time spent on the concepts of inquiry and problem solving.

BEST COPY AVAILABLE

Name	Address	City		Zip	Home Phone	School Name	School Address	School City		School Zip	School Phone	Fax Number	E-Mail Address
ee Morgan Sattelmeyer	3165 Ludlow Road	Shaker Heights	OH	44120	(216) 921-9155	Lomond Elementary	17917 Lomond Blvd	Shaker Hights	OH	44122	(216) 295-4050	216-295-4016	ralph@en.com
usan K. Robinson	227 Lakeview Drive	Millbury	OH	43447		Lemoyne Elementary School	4230 Fremont Pike	Lemoyne	OH	43725	(419) 837-2820	(419)372-7048	robinson@wbgu.bgsu.edu
enny L. Dilley	10271 Twin Oaks Drive	Cambridge	OH	43725	(614) 439-2040	Monroe Elemtary School	5000 St. Rt. 38	London	OH	43140	(614)432-6361	(614) 432-1086	CA_34@OMALP1.OMERES.OHIO.GOV
ristal A. Seagraves	5280 Center Street	Hilliard	OH	43026	(614) 777-0408	Monroe Elementary	5000 St. Rte. 38	London	OH	43140		(614) 873-4252	
am Styles	6600 Caldero Court	Dayton	OH	45415-1542	(513) [illegible]90-2908	Meadowdale Elementary	4448 Thompson Dr.	Dayton	OH	45416	(513)276-2104	(513)278-7806	kindypam@aol.com
ois S. Klamar	9079 Gettysburg Drive	Twinsburg	OH	44087	(216) 425-2674	Jamison CompuTech Center	13905 Harvard Avenue	Cleveland	OH	44105	(216) 295-0655		lklamer@wviz.org
ristine K. Dobbelaere	20486 Road 108	Oakwood	OH	45873	(419) 594-2206	Oakwood Elementary School					(419)594-3346	(419)594-2206 (call first)	krisd@bright.net
ary W. Cruser	Adams County Educational Service Center 141 Lloyd Road	West Union	OH	45693	(513) 544-5586	Peebles Elementary	P.O. Box 247	Peebles	OH	45660	(513) 544-3720		mcruser@scoda.ohio.gov
Michelle Kendrick	836 Leonard Avenue	Zanesville	OH	43701	(614) 452-3583	Frazeyburg Elementary	State Street	Frazeyburg	OH	43822			tr_6@omalph1.omeresa.ohio.gov

Name	Address	City		Zip	Home Phone	School Name	School Address	School City		School Zip	School Phone	Fax Number	E-Mail Address
Roy W. Gordon	22742 Holycross-Epps Road	Marysville	OH	43040	(513) 642-3588	Champaign County Educational Service Center	2380 S. Rt. 68, Box 269	Urbana	OH	43078	(513) 653-5296		rgordon@wviz.org
Kay A. Deitchel	c/o Scioto County ESC 411 Court St.	Portsmouth	OH	45662	(614) 353-8029	Scioto County Educational Service Center	411 Court Street, Room 105	Portsmouth	OH	45662	(614) 354-7761	(614) 353-1882	kdiechtel@scoca.ohio.gov
Debra Rucker	41 Keethler Drive North	Westerville	OH	43081	(614) 890-4575							(614)487-5380	drucker@magnus.asc.ohio-state.edu
Brenda F. Profit	1653 Glenmont Road	Cleveland Heights	OH	44118	(216) 932-7461	Jamison CompuTech Center	13905 Harvard Avenue	Cleveland	OH	44105	(216) 295-0655		
Mary B. Robinsonq	1185 Lectric Lane	Zanesville	OH	43701	(614) 452-1502	Perry Elementary	6975 East Pike	Zanesville	OH	43701	(614) 872-3436		
Bonniejean Meyer	10702 Notabene Drive	Parma Heights	OH	44130	(216) 884-0184	Tremont Elementary School	2409 W 10th St.	Cleveland	OH	44113			
Dawn Tufto	5461 Lakota Dr.	Westerville	OH	43081	614-882-9359	Taylor Rd. Elementary	8200 Taylor Rd.	Reynoldsburg	OH	43068			
Susan Dackin	8733 Taylor Woods	Reynoldsburg	OH	43068	614-863-1019	Taylor Rd. Elementary	8200 Taylor Rd.	Reynoldsburg	OH	43068			

SUMMER INSTITUTE ON SOFTWARE EVALUATION FOR K-4 MATHEMATICS AND SCIENCE EDUCATORS

This is a call for candidates to apply for a summer institute on software evaluation for K-4 mathematics and science. Participants in this institute will be learning how to use a protocol for evaluating software, will apply this protocol to evaluating a pool of software housed at the Eisenhower National Clearinghouse on Mathematics and Science Education (ENC), and will help us further refine and develop the instrument. The instrument that the participants will be using has been co-developed by faculty from the Reynoldsburg Public Schools and by staff from ENC and the Ohio Supercomputer Center.

There will be a total of 16 participants in this institute, who will meet at the Ramada Inn in Columbus from August 5-16. Participants will be compensated at a rate of $200/day and all travel, meal, and hotel costs will be covered. Travel to and from home over the weekend (Aug. 10-11) will also be reimbursed. (However, if an applicant wants to stay the weekend in the hotel, he or she will have to assume this cost.) Participants will also be supported in providing professional development and co-development experiences with educators in their own region over the course of the next school year. A shorter, follow-up summer institute is planned for next summer (Summer of '97).

We are seeking applicants from across Ohio to ensure a broad regional representation. They will be nominated by staff from the Regional Communities of Practice with this in mind. We are looking for applicants who possess the following five qualities:

- they should have extensive experience in teaching or working with teachers in a K-4 setting;
- they should be familiar with constructivist approaches to learning in general and hands-on, conceptually-oriented math and science instruction in particular;
- they should be knowledgeable about both the content and pedagogy of mathematics and science;
- they should be comfortable with using computers; and
- they should be prepared to assume leadership in organizing and conducting professional development in their own region.

Promising Practices in Mathematics and Science Education
Site Visit Validation Report

Title of Program ______________________________

School District ______________________________

School Name ______________________________

City ______________ State __________ Zip __________

Contact Person ______________________________

Title ______________________________

Telephone Number (____) ______________________________

We have verified the accuracy of the program descriptions as submitted on the nomination form and previously rated by a review team of content area experts. This program has been confirmed as a Promising Practice in Mathematics and Science Education.

__________ Yes
__________ No

Reviewers:

____

____

TO APPLY, CANDIDATES SHOULD: 1) WRITE A LETTER OF INTEREST INDICATING THE WAYS IN WHICH THEY POSSESS THE ABOVE FIVE QUALITIES; 2) ATTACH A RESUME; AND 3) INCLUDE ANY ARTIFACTS (LESSON PLANS, WORKSHOP DESCRIPTIONS, ARTICLES) THAT DOCUMENT ANY OF THE ABOVE FIVE ABILITIES.

THE APPLICATION PACKET SHOULD BE SENT TO:

THE OHIO SUPERCOMPUTER CENTER
1224 KINNEAR RD.
COLUMBUS, OHIO 43212
ATTN: TODD FENNIMORE

BY JUNE 7, 1996.

Any further questions or comments should be directed to Todd at 614-292-8985 or 614-294-4267.

☆U.S. GOVERNMENT PRINTING OFFICE: 1999 - 449-483

United States
Department of Education
Washington, DC 20208–5643

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (B)

ORAD 1999-1203-I

Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs

Working Papers

Volume Two

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs

Working Papers

Volume Two

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

Office of Reform Assistance and Dissemination
Peirce Hammond
Director

May 1998

Bound Edition March 1999

This publication is intended to promote the exchange of ideas among researchers and policymakers. The views expressed in it are part of ongoing research and analysis and, except for the official regulations, do not necessarily reflect the position of the U.S. Department of Education. Previously published documents such as articles from the *Educational Researcher* have been reproduced with permission of the authors and editors.

Note: In order to speed distribution of these working papers, we have reproduced original documents and their pagination.

Contact:
Susan Klein, Ed.D
(202) 219-2038
sue_klein@ed.gov

Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs
(Volumes One and Two)

Tab

VOLUME I.

Introductory Papers

Overview: Learning from Consumer-Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share the Best -- Susan Klein, Office of Educational Research and Improvement (OERI) A

Preface: Michael Scriven, Claremont Graduate University B

Current Status of the System of Expert Panels: Sharon Bobbitt and Susan Klein, OERI (April 1998) C

Incorporating Research Findings and Practice in Expert Panel Work: A Dialog Between Michael Scriven and Lois-ellin Datta (February 1997) D

II. Papers Examining other Review Activities to Identify the Best

A Synthesis and Integration of U.S. Evaluation Efforts to Identify Promising and Exemplary Educational Programs, Products and Practices -- Gary Borich, Univ. of Texas, Austin (July 1997) E

Examining Federal Approaches Outside the Department of Education to Identify and Disseminate the Best -- Karen Bogart, Anne Steinmann Institute (June 1997) F

Notes on the Federal Dissemination Experts Meeting: "Advising the Department of Education's Office of Educational Research and Improvement on its Design of a Systematic Consumer-oriented Evaluation System to Designate Promising and Exemplary Educational Products, Programs and Practices" Oct. 12, 1995, Academy for Educational Development, Washington, DC. -- Karen Bogart, Ann Steinmann Institute G

Standards and their Use in the Food and Drug Administration (FDA) -- Gerald (Jake) Barkdoll (Jan. 1996) H

Review of Foundation, Associations, and Non-profits Practices in Designating Promising and Exemplary Programs-- Janet Carter and Diane Schilder (July 1997) I

Learning from Consumer-Oriented Review Efforts in a Wide Variety of Education Organizations and Topic Areas -- John Luczak & Joan Ruskus, SRI International (July 1997) J

A Discussion of Some U.S. Evaluation Efforts for Programs and Resources in Mathematics and Science -- Carol Muscara, Computer Technology Services, Inc., Rockville, MD (Nov. 1996) K

Tab

VOLUME II

III. Papers on Schoolwide Models

Evaluation and Standards for Schoolwide Programs and Programs Conducted in Multiple Sites -- John H. Hollifield, Samuel C. Stringfield, and Rebecca Herman, Johns Hopkins Univ. (Jan. 1996) L

Changing the National Diffusion Network to Address Whole-School Reforms -- Susan Bodily and Thomas Glennan (Oct. 1996) M

IV. Legislation and Regulations

Key Parts of the 1994 OERI Authorization relating to Expert Panels N

Draft Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI) -- Designation of Exemplary and Promising Programs" Proposed Rule, June 3, 1996 O

Final Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI) -- Designation of Exemplary and Promising Programs, Nov. 1997 P

V. Conceptual and Planning Papers (in Chronological Order):

List of Papers and Publications Related to the System of Expert Panels Q

Developing and Operating Expert Panels: A Key Leadership and Coordination Role for OERI -- Susan Klein, OERI (Jan. 16, 1995) R

Leadership in Developing a Nation-wide "Findbest" Education R&D Evaluation System: A New OERI Responsibility -- Susan Klein, OERI (Jan. 29, 1996) S

Technical Assistance and the Creation of Educational Knowledge -- Brenda J. Turnbull, Policy Studies Associates (March 1996) T

OERI Leadership Efforts in Designing Approaches to Identify and Share Promising and Exemplary Products, Programs and Practices: Presession to the 1996 Annual Convention of the American Educational Research Association, New York City, April 8, 1996 Meeting notes by Allen Schmieder, OERI U

Responses to: Options in Discussing Distinctions Between Promising and Exemplary Products, Programs, Practices, Policies and Research Findings for a System of Expert Panels -- Susan Klein (Nov. 19, 1996) V

VOLUME II, Continued

	Tab
Expert Panels for Promising and Exemplary Innovations: A "Fine Idea From the Feds" by Peter Seidman, NCRVE Centerwork: National Center for Research in Vocational Education, Univ. of California, Berkeley (Winter 1997)	W
"Design Competitions: A Proposal for a New Federal Role in Educational Research and Development" by Robert E. Slavin, *Educational Researcher* (Jan./Feb. 1997)	X
"Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop, and Share Promising and Exemplary Products and Programs" by Susan S. Klein, *Educational Researcher* (Aug./Sept. 1997)	Y
The Evolving System of Expert Panels, A Presession on April 13, 1998 at the American Educational Research Association Conference, San Diego, Sponsored by the Office of Educational Research and Improvement, OERI, U.S. Department of Education. (Prepared by Chris Dwyer, Susan Klaiber and Peggy Simon, RMC)	Z
The Expert Panel System is a Reasonable Way to Identify Promising and Exemplary Programs --- A Response to Stanley Pogrow's Article, "What is an Exemplary Program and Why Should Anyone Care? A Reaction to Slavin and Klein." (Susan S. Klein, October 26, 1998).	AA

VOLUME II

Section III. Papers on Schoolwide Models

Evaluation and Standards for Schoolwide Programs and Programs Conducted in Multiple Sites -- John H. Hollifield, Samuel C. Stringfield, and Rebecca Herman, Johns Hopkins Univ. (Jan. 1996)	**L**
Changing the National Diffusion Network to Address Whole-School Reforms -- Susan Bodily and Thomas Glennan (Oct. 1996)	**M**

L

Evaluation and Standards for Schoolwide Programs and Programs Conducted in Multiple Sites

John H. Hollifield, Samuel C. Stringfield, and Rebecca Herman
Johns Hopkins University

This paper discusses issues involved in the development of appropriate standards for designating schoolwide programs and programs conducted in multiple sites as promising or exemplary. The scope and complexity of these programs poses many problems for their evaluation and for how they might be described as either promising or exemplary.

In Part I of this paper, we describe these programs and examine evaluations that have been conducted on their implementation and effects, including the recent Special Strategies Study, a comprehensive evaluation of the actual effects of many of these programs being implemented in local sites. Further in-depth descriptions and reviews of evaluations are included in Herman and Stringfield (1994) and Stringfield et al (1994). In Part II, we discuss how our experience with these evaluations provides (or doesn't provide) a base for designating various programs as exemplary and promising. This part includes discussion of issues concerning the strength of our evaluations, variations in the philosophies and components of the programs as they influence issues of implementation and replication, and the measurement of student achievement and learning. In Part III, we summarize the implications of these evaluations for the design of standards for designating promising and exemplary schoolwide and multi-site education programs, and suggest some wording for the standards.

PART I: DESCRIPTION OF PROGRAMS AND REVIEW OF EVALUATIONS

Title I (Chapter 1) Schoolwide Programs

Title I Schoolwide Programs -- Models

Schoolwide compensatory education projects (SWPs) provide an opportunity for high-poverty schools to use Title I (Chapter 1) funds to serve all students. Prior to 1988, schools were permitted to serve all students in schools in which at least 75 percent of students qualified for free lunch, but only if the district provided substantial matching funds. Since high-poverty schools frequently are located in high-poverty districts, the matching funds criterion made schoolwide projects impractical for most districts. However, the Hawkins-Stafford amendments of 1988 dropped the matching funds requirement, making the schoolwide option far more accessible to high-poverty schools. As a result, there was a movement toward schoolwide projects in many districts: from 199 schools in 1988-89 to 621 in 1989-90 (Turnbull, Zeldin, and Cain, 1990). Schenck (in U. S. Department of Education, 1993) noted that more than 2,000 out of 9,000 eligible schools operated schoolwide projects.

P.L. 103-382, the 1994 Improving America's Schools Act (effective July 1, 1995), provides even greater freedom to schools and districts in how they spend their Title I funds and lowers their minimum poverty requirements for implementing schoolwide programs. It is estimated that due to this Act, with its new freedoms, up to 90 percent of the now estimated 20,000 schools eligible to implement schoolwide programs will do so starting in the 1996-97 school year (Hoff, 1994).

Schoolwide projects have traditionally reduced class size, reduced or eliminated pull-out instruction, increased staff development, and made Chapter 1 materials available to all students (Stringfield, Winfield, Millsap, Puma, Gamse, and Randall, 1992). Class size reduction has been the most common response to schoolwide opportunity (Winfield, Hawkins, and Stringfield, 1992; Committee on Education and Labor, 1990; Turnbull et al., 1990). Seventy-nine percent of schoolwide projects included a reduction in class size (Turnbull et al., 1990). In some cases, the entire program consisted of reducing class size with no accompanying staff development to promote new strategies.

The second most popular schoolwide strategy has been the use of "effective school" programs (62 percent). Some of these programs are discussed in detail below. Seven percent of schoolwide projects used the funds to extend the school day (Turnbull et al., 1990).

Pechman and Fiester (1996) describe what seems to be emerging as a "framework" for successful schoolwide programs: an agreed-upon vision; time and resources for planning and program design; a well-defined management and organizational structure; a clear focus on academics; continuing schoolwide professional development; a commitment to cultural inclusiveness; parent and community involvement; and an accountability orientation. This framework corresponds well to the components of schoolwide programs that are required under the new Title I, as described by LeTendre (1996): a comprehensive needs assessment of the entire school in relation to state content and performance standards; schoolwide reform strategies that are based on effective means of improving the achievement of all children; effective instructional strategies; professional development for teachers, aides, other staff, and parents; strategies to increase parent involvement; strategies to assist preschool children in transition from early childhood programs; and activities to ensure that students who experience difficulty in mastering state standards will be provided timely, effective assistance.

Title I Schoolwide Programs -- Evaluations

Pechman and Fiester (1996) point out the difficulties involved in evaluating the effectiveness of schoolwide programs: the general lack of longitudinal data, the tremendous variation in actual programs across schools, and the fact that a schoolwide program is essentially a management strategy, not an intervention model, and implementation hinges on the strengths of individual leaders and teachers. "As a result," note Pechman and Fiester, "findings to date about schoolwide outcomes are mixed and inconclusive."

Evaluations of programs that have initiated class-size reductions reflect the mixed and inconclusive effects. Doss and Holley (1982) evaluated schoolwide projects in Austin, Texas in which Chapter 1 and other funds were used to reduce class size from 25-1 to 15-1. Initial results were encouraging, but a five-year follow-up (Christener, 1987) found no achievement differences between the smaller classes and matched classes in other schools. Other research on substantial reduction in class size has found small to moderate effects averaging only .17 standard deviations in the early grades and less than that beyond grade three (Slavin, Karweit, and Wasik, 1994).

Other research indicates that schoolwide programs have had positive, if mixed, effects. Davidoff and Pierson (1991), Winfield, Hawkins, and Stringfield (1991), and Winfield and Hawkins (1993) have conducted evaluations of Schoolwide Program effects in Philadelphia, a district that adopted the schoolwide philosophy early and extensively. Davidoff and Pierson compared schoolwide program effects to traditional Chapter 1 program effects over a five-year period (1985-90), and found consistently positive effects on reading and math (NCE scores), student grades, and student attendance. Winfield, Hawkins, and Stringfield, in case studies of six schools, found increased (but inconsistent) reading and math achievement. Winfield and Hawkins, in longitudinal evaluations of 40 elementary schoolwide programs compared to 20 elementary schools with similar levels of poverty, found very mixed results -- achievement results fluctuated by grade level and were heavily influenced by age, gender, and race/ethnicity of the students.

Although the overall evaluation of the effects of schoolwide programs as a whole is very problematic, evaluation of the effects of a specific schoolwide reform implemented in a specific school is more straightforward. Numerous case studies that document effectiveness have been produced by evaluators (e.g., Pechman and Fiester, 1994, 1996; Leighton, Hightower and Wrigley, 1995), including those in the recent Special Strategies Study. Also, many individual schools have submitted documents to the Chapter 1 National Recognition Program that carefully describe the elements of their programs, discuss their processes of implementation, and present data on the effects of their schoolwide programs in many areas, including student achievement. Many case studies -- some more extensive and convincing than others -- are available on the PRC Inc.'s Chapter 1 Bulletin Board System and in the conference notes compiled for each annual National Chapter 1 Schoolwide Projects Conference (U. S. Department of Education, 1992, 1994).

Special Strategies Evaluation of Schoolwide Projects

The Special Strategies Study examined four sites using schoolwide strategies. Two sites are in an urban district which targeted schoolwide programs across the district. The other two sites, in rural districts, chose schoolwide strategies at the school but not district level.

Several approaches were emphasized at multiple study sites. Three of the sites stressed whole language instruction, either alone or in conjunction with strategies such as phonics. Interdisciplinary thematic lessons were common at three sites. For example, one urban school operated an interdisciplinary creative project for all students, the Communication Arts Network

Program. A rural school was moving towards an integrated reading-writing-arts program. Several schools provided computer-assisted instruction to all students. Computer-assisted instruction tends to occur frequently in schoolwide projects; this is one example of Chapter 1 materials becoming available to a broader group of students. The two urban schools used a variation of the Madeline Hunter instructional model.

Programs Conducted in Multiple Sites

The programs examined in this section include the School Development Program, the Success for All Program, the Paideia Program, the Essential Schools Program, the Accelerated Schools Program, and the programs being developed under the auspices of the New American Schools Development Corporation (NASDC).

Comer School Development Program – the Model

James Comer has developed and implemented a schoolwide restructuring project, the School Development Program (SDP), to address the needs of the whole child. This program emphasizes school-based mental health services, parent involvement, and teacher involvement in restructuring the school's programs. Nine components--three guiding principles, three mechanisms for development, and three operations--are considered essential:

- "no fault" decision making (principle),
- consensus decision making (principle),
- collaboration (principle),
- parent involvement program (mechanism),
- school planning management team (mechanism),
- mental health team (mechanism),
- a comprehensive school plan (operation),
- staff development (operation), and
- assessment and modification (operation).

Central to this model is a school management and governance team composed of the principal, teachers, parents, a mental health specialist, and support staff. This team develops the school's "master plan." Although on-site players determine the specific shape of the program, the instructional program includes a Focus Program, a small-group pull-out tutorial provided at least three times per week to students who are at least a year behind grade level, and a Discovery Room to entice or draw out troubled learners (King, 1994). The School Development Program is essentially "content-free" and, in principle, could be adapted to diverse local curricula.

The SDP is now operating in over 150 schools in 14 school districts. It was first implemented in two New Haven elementary schools by the Yale University Child Study Center in collaboration with the New Haven School System.

Comer School Development Program -- Evaluations

School- and district-level evaluations have found SDP to increase student achievement in comparison to similar schools (Comer, 1988; Joyner, 1990). SDP schools had significantly greater percentile gains on the California Achievement Test compared to their district as a whole (Haynes and Comer, 1991). Based on data on randomly selected students in matched schools, students in SDP schools achieved significantly higher averages in mathematics and overall grades and in reading, mathematics, and language scores on the California Achievement Test (Boger, 1989; Cauce, Comer and Schwartz, 1987; Haynes, Comer and Hamilton-Lee, 1988b, 1989). In a follow-up study using course grades and percentile scores on the Metropolitan Achievement Test (MAT), sixth grade SDP students scored higher on all achievement measures, but only significantly higher on some of the subtests of the MAT (Haynes and Comer, 1990; Comer, 1988). In the same study, eighth grade SDP students again scored consistently, but not significantly, higher than non-SDP students.

Because this model focuses on children's personal, social, and academic growth, non-academic outcomes are also relevant. SDP students tend to have better self-concept compared to control groups (Boger, 1989; Haynes and Comer, 1990; Haynes and Emmons, 1990a, b, c; Joyner, 1990). Research shows mixed effects of the program on attendance. In one study, SDP students had significantly better attendance compared to non-SDP students (Haynes, Comer and Hamilton-Lee, 1988, 1989). In a second study, SPD and control students had similar attendance rates (Joyner, 1990). Aggregated data analysis over four years in a Michigan district shows that fully-implemented SDP schools achieved significant declines in suspension and absence days and amount of corporal punishment compared to non-SDP schools (Haynes, Comer, and Hamilton-Lee, 1988a).

Special Strategies Study Evaluations of the Comer School Development Model

Observations and interviews suggest that the principal's approach and teachers' support for the program greatly influence its success. These are factors that can be addressed in future implementations. For example, a well trained facilitator, such as those used at some of the replication sites, might ease political differences and train staff to work cooperatively. One key to success seems to be a conscious decision by the principal and teachers to share responsibility for re-creating their school.

Success for All Program -- the Model

Success for All is a schoolwide restructuring program designed to see that students begin with success in the early grades and then maintain success through the elementary years. Slavin et al. (1992) declare that it combines many of the effective approaches identified in prior research, including the following:

- Research-based preschool;
- Whole-day kindergarten;
- Beginning reading programs for 90 minutes per day, integrating the phonics and whole language approaches in a set sequence of story, discussion, new vocabulary, oral

language production, comprehension, and story structure;

- Homogenous reading groups, reassessed and regrouped every eight weeks;
- One-to-one tutoring, integrating the regular curriculum, for students experiencing difficulties in reading in grades one through three, with priority given to first grade students;
- Cooperative learning approaches to intermediate reading, writing/language arts, and mathematics;
- Family support services to increase parent involvement and to remedy home-based problems (i.e., poor attendance) that interfere with learning; and
- A part- or full-time project facilitator to coordinate the many program elements and provide training and technical assistance.

The goal of keeping students performing on grade level involves minimizing retention and avoiding dependence on long-term special services such as special education classes. Difficulties in learning are addressed, as much as possible, within the context of the regular classroom.

Success for All -- Evaluations

Longitudinal studies, using matched control students in matched schools, indicate that Success for All improves student achievement. Success for All students in five Baltimore schools had significantly higher reading achievement compared to matched controls and surpassed control students by three to seven months in average grade equivalents (Madden, Slavin, Karweit, Dolan, and Wasik, 1993; Slavin, Madden, Karweit, Dolan, and Wasik, 1992). Students who scored in the lowest 25 percent on the pretest--the most at-risk students--showed the most improvement (Slavin et al., 1992). These improvements lasted through most elementary grades in programs which had implemented Success for All for at least five years (Slavin, Madden, Dolan, Wasik, Ross, and Smith, 1994).

Research has found that Success for All, compared to other Chapter 1 schoolwide projects, reduces retention and special education assignments and increases attendance (Madden et al., 1993; Slavin, Madden, Dolan, Wasik, Ross, and Smith, 1993). Reduction in retention is an element of the program rather than an outcome; however, passing marginal students while providing them with additional academic support seems to have promoted greater gains compared to controls who were retained. Reduction in special education assignment is in part policy and in part outcome of the program. In traditional schools, students often are identified as learning disabled as a result of reading far below grade level. A premise of the program is that learning problems such as below-grade level reading skills should be resolved outside of special education. In Success for All schools, only 3.9 percent of third graders were reading two years or more below grade level, compared to 11.7 percent of third graders in control schools (Slavin et al., 1992). A study comparing Success for All with Reading Recovery showed complementary strengths of the two programs: Reading Recovery provided a focused, intensive intervention for a small group of at-risk students while Success for All reached a broader group of at-risk students (Ross, Smith, Casey, and Slavin, 1994). Both improved the reading of tutored students, but Success for All also improved the reading performance of students who were not tutored.

Evaluations of other Success for All programs, including a Philadelphia program for language minority children, reiterate the findings in Baltimore (see Ross and Smith, 1992; Slavin et al., 1993; Slavin and Madden, 1994; Slavin and Yampolsky, 1992; Wasik and Slavin, 1993). A multi-site replicated experiment, pooling effects of 34 first grade cohorts in 13 schools, found that the difference between Success for All students and controls increased from first, to second, to third grade (Slavin et al., 1993). According to these studies, Success for All students learned significantly more than matched control students and continued to perform at a higher level. Recent evaluations reporting on the longitudinal study of Success for All effects from 1988 through 1994 of school year 1993-94 In September 1994, Success for All had been implemented in over 140 elementary schools in 56 districts in 20 states (Slavin et al., 1994). Successful replication suggests that the program can work in diverse circumstances and does not rely on a particular school structure or principal (Slavin et al., 1993).

Special Strategies Study Evaluations of Success for All

Of the programs in Special Strategies, Success for All has the strongest external research base. Success for All seems to help children learn to read and to enjoy reading. One strength of the program is that the key elements -- regrouping, assessment, individualization, remediation -- are systemic; the structure of the program, rather than the personalities implementing it, are designed to maximize the probability of academic success.

Paideia Program -- the Model

Paideia, the education model developed by Mortimer Adler (1984) in Paideia Proposal: An Educational Manifesto, focuses on improving the quality of education for all students, regardless of background. A central concept of the Paideia philosophy is that high academic achievement is expected of all students. Adler believes that low or high achievement is not a valid indicator of ability and should not be used to sort students into classes. The heterogeneous classes use original sources, but are not limited to "great books."

The goals of the Paideia program are "acquisition of knowledge," "development of intellectual skills," and "enlarged understanding of ideas and values" (Adler, 1984: p.8). These goals can be addressed through three instructional approaches, known as "columns":

- Didactic instruction. Teacher lectures provide opportunities for "acquisition of knowledge."
- Coaching. The students work at their own level and pace with one-on-one instruction from the teacher. In practice, peer tutoring and computer-assisted instruction are considered elements of coaching. "Development of intellectual skills" should be the focus of coached work.
- Small-group seminars. The teacher functions as an instructional facilitator rather than storehouse of knowledge, using the Socratic method of questioning to explore issues. This approach is appropriate for "enlarged understanding of ideas and values."

Typically, schoolwide restructuring is necessary to fully implement all three instructional approaches. For example, seminars often require longer than average class periods and coaching calls for smaller classes during parts of the day.

Paideia Program -- Evaluations

A 30-week study of an instructional model based upon Goodlad's community of learners idea, Anderson's quality circle management, and Adler's instructional model offers the opportunity to evaluate the Paideia ideas (Lumpkins, Parker, and Hall, 1991). A team of four full-time teachers and 120 fourth and fifth grade students participated in an experimental learning program. Instruction was based on elements of the Paideia model including didactic instruction, small-group seminars, and coaching. Low achieving students in both grades made significant gains in mathematics concepts and skills.

A 1987-88 evaluation of the Chicago program describes promising effects. The Paideia program in Chicago reached more than 5,000 students in grades K through 12 in 23 schools (Wallace, 1993). On average, students spent 12 hours per week in Paideia instruction--nine hours for didactic instruction, two hours for Socratic seminar, and one hour for coaching. A teacher coordinated the program at each school. Participating teachers had extensive opportunities for training in the Paideia approach, the Socratic method of questioning, drawing out reticent students, and controlling dominant students.

Evaluators of the Chicago program found that, compared to non-Paideia students at the same high schools, Paideia students had higher average daily attendance (84 and 91 percent versus 78 and 85 percent). On the Tests of Achievement and Proficiency, 33 to 66 percent of non-Paideia students in Paideia schools scored in the bottom quartile in reading comprehension, math problem-solving, and science compared to 11 percent of Paideia students. Citywide, Paideia elementary students had lower failure rates in English, math, or science (3.6 percent versus 4.3 percent). Also, students in the Paideia program seemed to express and support their ideas better than comparison students, based on scored writing samples. Finally, responses of students, teachers, and parents on the effectiveness of their program and their reactions to the program were extremely positive.

Special Strategies Study Evaluation of Paideia

Four Paideia sites were reviewed in this study: two primary and two replication sites. One primary and one replication site had been using the Paideia program since 1984. The other primary site began implementation in 1989. The final site, a K-12 school developed around the Paideia program, opened in 1986.

Only one of the four study sites had consistently implemented a three column Paideia program. The two primary sites appeared to have strong seminar instruction but less well-developed didactic instruction and coaching. One of the replication sites drifted from all three strategies for some time but returned under the guidance of a new principal.

Of the three instructional strategies, seminars were the most consistently implemented. Weekly seminars lasted from forty minutes at one site to two and one-half hours at another. Students at three sites prepared for the seminars by prereading assignments and preparing questions. Although the seminars themselves occupied a small portion of instructional time, teachers adopted many of the questioning strategies for regular classes.

Both primary sites were working towards a three-column Paideia program. One focused on improving didactic instruction. However, their chosen strategy, DISTAR, more closely resembled Adler's coaching. The other primary site de-emphasized coaching for approximately one year before formally reinstating it.

Teachers in all of the Paideia program sites visited reported improvements in students' abilities to think critically and express themselves clearly.

Unfortunately, the flexibility of the program appears to have contributed to uneven implementation at most sites. Many administrators might be reluctant to implement Paideia because of the program's abstract nature and the considerable investments required to make it practical. It is possible that extensive teacher training may compensate for teachers' varied initial experiences and skills. The program calls for committed staff, a well trained coordinator, and the consistent, multi-year support of the implementing school's administration.

Coalition of Essential Schools -- the Model

The Coalition of Essential Schools (CES) was founded by Theodore Sizer in 1984. Sizer developed the basic philosophy through a two-year study of American secondary schools and, in response to practitioner enthusiasm, promoted the ideas as a basis for school restructuring. The initial focus of CES was on the "triangle of learning" between teacher, student, and subject matter; the model has grown from classroom to school-based change as developers recognized that effective reform requires more fundamental changes in the school. Nine Common principles support this focus:

- Schools should have an intellectual focus;
- School goals should be simple;
- Schools should have universal goals that apply to all students;
- Teaching and learning should be personalized;
- The governing metaphor should be student-as-worker, teacher-as-coach;
- Diplomas should be awarded upon demonstration of mastery;
- The tone of the school should stress unanxious high expectations;
- Principals and teachers should view themselves as generalists first and as specialists second; and
- Teacher load should be 80 or fewer students and per pupil cost should not exceed traditional school costs by 10 percent.

The principles are deliberately ambiguous to encourage schools to develop their own strategies within their particular context. Sizer and his staff actively avoid promoting a "boilerplate" model.

In 1988, CES and the Educational Commission of the States (ECS) created "Re:Learning," a joint reform effort. CES continued to focus on site-based reform while ECS worked on state-level policy reform and recruitment. There are now over 700 schools using CES ideas through the Coalition, Re:Learning, and related networks (Riggs, 1994).

Coalition of Essential Schools Evaluations

CES schools are difficult to evaluate, first, because each implementation is intended to be unique and second, because the goals of the program are viewed by the developer as not being at all aligned with traditional norm-referenced tests. The Coalition refuses to recommend a specific "model" because no single model could address each school's different needs and assets. However, there are some efforts to determine whether Sizer's approach affects educational outcomes.

CES has launched a nine-year longitudinal study, "Taking Stock," which will follow 50 to 75 program students through high school and five years beyond (Cushman, 1991). Consistent with the CES approach, in which the school determines the goals and shape of the program, the evaluation will use measures other than nationally normed tests. For example, course grades and attendance, suspension, and graduation rates could indicate program success. The pilot study for "Taking Stock" found lower dropout rates and higher attendance in Essential Schools compared to district rates or CES school pre-implementation rates. Based on achievement test scores, academic performance in some Essential Schools has also been reported to have improved (Cushman, 1991). Fewer discipline cases were CES students and, in some cases, more CES students went on to higher education. These findings are based on a selective review of existing records and, although encouraging, should be interpreted with caution.

A qualitative review of a CES school and an outcome-based education (OBE) school found that both strategies caused teachers to internalize student-centered approaches to teaching and assessment (Desmond, 1992). CES teachers focused more on performance assessment and collaborated professionally on assessment more often compared to OBE teachers.

The amorphous nature of CES not only complicates evaluation, but also confounds implementation. Hampel (in press) found specific difficulties related to implementing CES centered on divisions among staff, a common finding in much of the research on CES implementation. Part of understanding the phenomenon of staff divisiveness is examining how the divisions arise. A study of CES schools in South Carolina (Shirley and Anderson, 1994) found that teachers and administrators tended to support Re:Learning, but their conceptions of the project differed. Thus, they were trying to implement essentially different programs. Educators felt that some of the nine essential principles were more important than others, exacerbating the difference in perceptions of the program. Further, diffuse leadership failed to bring the actors together to a common program and goals. Individually, a "critical mass" of teachers in each school use the CES principles in operating their classes, but a cohesive CES program was not established.

Chicago's experiment with CES was unable to overcome implementation difficulties. Most of the city's eleven CES schools dropped in attendance and graduation rates and standardized test scores and increased in dropout rates after adopting CES (Sikorski, Wallace, Stariha, and Rankin, 1993). These poor outcomes were attributed to inconsistent implementation and emphasis on governance rather than academic issues.

The School Ethnography Project, a five-year study of eight charter Coalition schools, found that political tension tended to impede the restructuring effort (Muncey, 1994; Muncey and McQuillan, 1993). Proposed and enacted reforms created and exacerbated political tensions.

Not all schools were stymied by implementation difficulties, however. One exemplary Coalition school, Broadmoor Junior High School in Pekin, Illinois, was the subject of an intense, three-year case study (Prestine, 1993). The project began in Spring, 1989. Poor results during the first year of implementation were attributed to partial, rather than full, implementation of the nine principles and inadequate training and planning. A reconceptualized reform effort the following year was much more successful, according to teachers, administrators, and parents. Implementation was much smoother and resulted in a lower student suspension rate. Changes in ideas and attitudes, more sophisticated restructuring, a flexible approach, support of the superintendent, and systematic application of all nine principals contributed to the success of the restructuring.

Prestine's larger qualitative study (1994) suggests that even the most enthusiastic reformers may flag trying to implement their CES project. Because of their limited resources, schools often must choose where to focus their time and materials; pragmatic concerns such as new state mandates repeatedly win over philosophical concerns such as essential school restructuring. Personnel turnover, a constant in most schools, also undermines the stability of implementation. Teachers and administrators sometimes expend all of their effort on designing a specific program that reflects the CES principles and have little time, energy, or enthusiasm for implementing the program. Prestine suggests that not having a program model helps in developing school-specific change, but the school's strategy, once developed, should be formalized into standard operating procedures if it is to last.

In response to implementation difficulties, the Coalition has begun to recommend certain approaches to reform. For example, many schools that attempted to ease CES into the school using a school-within-a-school structure failed to make the next step to schoolwide implementation (Lessons from the Trenches, 1990). Staff split into factions which became more entrenched the longer it took to expand CES. Consequently, the Coalition has begun to recommend that schools operationalize their program more quickly (between two and seven years) and broadly (Prestine and Bowen, 1993). The Coalition identified four benchmarks of change--substantial agreement, observable change, all-school participation, and systemic leadership--and encouraged schools to evaluate their progress along these dimensions (Prestine and Bowen, 1993).

Special Strategies Study Evaluations of the Coalition of Essential Schools
The five sites in the Special Strategies studies illustrate the diverse approaches to Sizer's Coalition of Essential Schools. These CES sites were instigated in the 1980's at the state, district and school levels. The state Governor supported the innovation in one case and the state Department of Education in another. Two superintendents, four principals, and several teachers were instrumental in adopting CES.

None of the sites began as a schoolwide project. In two cases, the program began as a school-within-a-school; a third program evolved into a school-within-a-school by its second year. Two schools were expanding their programs to encompass the entire school; the other maintained a CES school-within-a-school alongside two other models. Two sites did not implement CES in a discrete structure. In these sites, specific teachers and teams of teachers used CES strategies. As the Coalition grew in popularity, more grades and teachers were incorporated.

The school's strategy for making the program schoolwide probably has some effect on staff acceptance of CES. Sites which expand the program one grade at a time seemed relatively more successful at minimizing staff friction and maintaining equivalent levels of staff expertise across the grade level. At the site which allowed teachers to individually implement specific CES projects, firm, quiet opposition to CES solidified and support for CES eventually dwindled.

Some common elements arose among the cases. Block scheduling, especially of English and social studies and interdisciplinary units, reflects the CES philosophy of focusing on and integrating basic subjects. The student-as-worker often was realized through project-oriented and group work. The teacher-as-facilitator could be seen in open-ended questions as opposed to lectures, and in the teacher team focus on individual students.

Accelerated Schools -- the Model
The primary goal of the Accelerated Schools model is to accelerate, rather than remediate, the learning of at-risk students so that they perform at levels appropriate to their age group. The model has three central principles:

1. Unity of purpose: the development and pursuit of a common vision among parents, teachers, support staff, students, and building administrators of a school as well as central office administrators, which will work for all students.

2. School-site decisions and shared responsibilities: the empowerment of key participants to make important decisions at the school level and in the home to improve the education of all students.

3. Building on strengths: utilizing all the learning resources that students, parents, school staff, and communities can bring to the educational endeavor -- using everyone as "experts."

Elements of Accelerated Schools include powerful learning as a basic premise, empowerment of all stakeholders, ownership of the Unity of Purpose, change that builds on the strengths of all, a governance structure that provides the basis of collaboration through the inquiry process, and a grounding in the values of acceleration, equity, participation, cummunication, community, risk taking, reflection, experimentation, and trust.

Accelerated Schools -- Evaluations

Evaluations of the effects of the Accelerated School on student achievement, school processes, and other outcomes exist primarily as case studies. These can be found in issues of the Accelerated Schools newsletter and in *Accomplishments of Accelerated Schools*. Individual schools using the Accelerated Schools model have been recognized for "excellence" by state recognition programs and by the national Blue Ribbon Schools Program.

Evaluation outcomes such as the following are reported:

> ... reducing the mobility rate from 30 percent to 23 percent, improving student behavior with only 34 total days of suspension as compared to 103..., increasing the level of parent involvement, and improving test scores of sixth-graders on the Comprehensive Test of Basic Skills in all three areas. In schoolwide surveys, students and staff report higher levels of self-esteem and enjoyment from school than ever before. (Edison Accelerated School, CBBS Bulletin Board)

New American Schools Development Corporation (NASDC) -- Designs

The New American Schools Development Corporation is a corporate sponsored organization set up in 1991 to help develop high performance schools by supporting the development and then scaling up of nine school improvement designs. The Education Commission of the States (ECS) is a primary partner of NASDC. In Phase 1 of this effort, the nine designs were developed and implemented in collaboration with schools. Phase 2, the demonstration and test stage, is now in process, with the designs being implemented in jurisdictions nationwide.

The nine designs are: Atlas Communities, The Audrey Cohen College System of Education, Community Learning Centers, Co-NECT Schools, Expeditionary Learning Outward Bound, Los Angeles Learning Centers, Modern Red Schoolhouse, National Alliance for Restructuring Education, and Roots and Wings.

Each New American Schools' design is built on a set of principles that describes a philosophy about what best serves students (Education Commission of the States, 1994). The nine designs have attributes in common: a focus on the whole school rather than on one or two elements; a focus on active learning, student engagement, and up-to-date technology and instruction; an emphasis on student groupwork in classrooms and communities on real-life projects; high academic standards; professional and staff development; and parent and community involvement. The designs measure achievement through standardized tests, portfolios, public displays, and other means.

New American Schools Development Corporation (NASDC) – Evaluations
Evaluations of outcomes for the NASDC designs have not yet been conducted, as most are early in the demonstration cycle. Some of the designs build upon previous reform efforts that have some evidence of positive outcomes -- for example, Roots and Wings builds upon the well-documented effectiveness of the Success for All Program. A formative assessment of the demonstration and test stage by RAND (Bodilly, in press) provides a progress report that examines the differences among the designs and the design teams, how those differences affected their progress, insights on useful implementation strategies for reform, and the contribution of the NASDC designs to reform thus far.

Part II: LESSONS LEARNED ABOUT EVALUATING BROAD SCALE PROGRAMS

Our review of schoolwide and multiple-site programs and their evaluations provides a basis of discussion about standards for designating programs as exemplary and promising. The main issues for discussion include the strength of the evaluations, variations in the programs and how those variations affect implementation, adoption, adaptation, and development, and the measurement of student achievement and learning. We discuss these issues separately but they impact heavily on one another.

Strength of Evaluations

The quality of evaluations of schoolwide and multiple-site programs has been extremely uneven. Only one program (Success for All) has undergone rigorous evaluation, using appropriate student achievement measures, a control group design, and extensive evaluations in replication sites. This program has shown strong effects on achievement in reading, improved attendance, and decreased retention and participation in special education. The School Development Program evaluations have provided strong evidence of increased student achievement in school- and district-level applications, some evidence of improved student self-concept, and, in fully implemented sites, decreases in suspension and absenteeism. In the case of these two programs, similar findings across studies suggest a reliable research base and lend credibility to the findings.

The research literature includes both quantitative and case study evaluations. Chapter 1 Schoolwide Programs have been evaluated against Chapter 1 schools using traditional programs, students' previous achievement levels on norm-referenced tests, and progress compared to district norms. Case study evaluations have been conducted of many specific schoolwide programs (including Accelerated Schools programs), and are the preferred method of evaluation for the "Taking Stock" initiative being carried out by the Coalition of Essential Schools. These case studies, when they include convincing achievement measures, can provide evidence of effectiveness in the particular schools being studied.

Numerous examples exist of individual schools in which the Chapter 1 schoolwide program is succeeding in improving student achievement, but these numerous examples represent probably less than 10 percent of all programs being implemented. The bias in published research is toward reporting successful cases. Thus, collectively, the evaluations of Chapter 1 schoolwide programs are far from conclusive.

Some programs have resisted evaluation. The Coalition of Essential Schools, from its inception, has argued that its outcomes should not be measured in traditional ways. Also, many evaluations have been stymied by implementation problems. The programs of the New American Schools Development Corporation are just now beginning to undergo evaluations of their effects, and these evaluations are seriously complicated in many cases by implementation issues..

The experience of the Special Strategies Study sums up the difficulties involved in conducting strong and rigorous evaluations of schoolwide and multiple-site programs. The study examined 10 programs in nominated sites and in replication sites. In an ethnographic component, it documented "whole school days" at the various sites, which involves following a student through an entire day in order to see if the program is actually making things different for that student. More often than not, little difference was observed -- the school and the classes in school and the relationships and interactions among students, teachers, and others varied sometimes but not often for the individual student. Usually, this was due to implementation failure -- even at sites nominated by the developers of these programs, components of the programs were either missing or had only a weak presence. If a program is not implemented, it can hardly be credited with creating any positive effects, or held solely responsible for a lack of positive results.

Nonetheless, the overall conclusion of the Special Strategies study is positive. The study finds that a more sophisticated, more mature field of "promising programs" exists in the 1990s than at any previous point in our history. Most of the programs studied have at least some research support to indicate that they may have positive effects on students. Some of the programs have substantial, sophisticated research support, both for their components and their over-all effects. Also, the Special Strategies study indicates that widespread implementation of these programs is possible -- the program successes, where they existed, were not limited to one or two sites nation-wide, or to sites near the developers, or to sites in one region of the country. To the extent that the programs "work," they can be made to work in often diverse contexts.

However, the Special Strategies Study also emphasizes that none of these programs are "teacher-proof, school-proof, or district-proof." District fiscal crises killed program implementations, almost regardless of the particular program. A faculty member's refusal to pay more than lip-service to any promising program, if tolerated by the school and district, affects the level of implementation of the entire program. A lack of long-term, program-specific staff development left one or more sites of several programs permanently "malnourished." Failure to build in systems and time to coordinate among faculty, even among faculty striving to successfully implement a program, severely limited implementation of promising programs. The fact that many of these programs *can* work is not evidence that one or more *will* work at a specific site with a particular team of teachers and administrators. Considerable care and support are needed for the successful implementation of any of these programs.

Program Variation

There are multiple issues of program variation to be resolved in designing standards for designating schoolwide and multiple site programs as exemplary and promising.

One major issue is the question of program replication in other sites. Our examination of the evaluations of schoolwide and multiple site programs finds wide disparities in the replicability of these programs. In general, some programs are school-based and specify procedures, practices, and materials that can actually be replicated in other schools, and staff development is focused on assisting schools in applying those specific procedures, practices, and materials -- thus the program when implemented in one school looks essentially the same as in other schools, and the outcomes are more assured. Other programs are school-based but promote more general guidelines and principles, and staff development focuses on assisting school personnel in working together, developing goals and objectives, and developing their own school-based set of procedures, practices, and materials. Thus the program implemented in one school may look very different from the program implemented in other schools, and the outcomes are less assured.

A good way to examine this difference in replicability is by comparing the "new" Title I schoolwide programs to the Success for All program. Title I schoolwide programs, as described in the new IASA, must have the following characteristics: a comprehensive needs assessment of the entire school in relation to state content and performance standards; schoolwide reform strategies that are based on effective means of improving the achievement of all children; effective instructional strategies; professional development for teachers, aides, other staff, and parents; strategies to increase parent involvement; strategies to assist preschool children in transition from early childhood programs, and activities to ensure that students who experience difficulty in mastering state standards will be provided timely, effective assistance. These constitute a specific set of guidelines and principles to be used by schools to develop an effective schoolwide program, and our review of evaluations finds that some schools succeed in doing so. But the components of the program at any one school may differ greatly from the components in another -- e. g., the effective instructional strategies used in schoolwide programs range from school-developed emphases on constructivist practices to whole host of proven effective practices (some more proven than others) -- Reading Recovery, varieties of cooperative learning, HOTS, CCC and Josten's computer programs, Writing to Read, Writing to Write, and so on. Pechman and Fiester (1996) note that "the most effective Title I schoolwide programs adapt the practices that researchers and practitioners have proved successful in the past." These practices vary tremendously from school to school and even within a school -- one schoolwide program

(named a 1993-94 National Blue Ribbon School) proudly claims a total of 41 major innovative programs operating in its classrooms (Kentucky Department of Education, 1995).

We can compare this variety in program implementation and lack of replicability with the Success for All program. Success for All specifies program elements which are implemented according to specification in schools which replicate the program. Variations occur as schools adapt the program, but Success for All remains recognizable as Success for All. What becomes immediately apparent in this comparison is that the Success for All program itself is a schoolwide program that has specified and operationalized the general characteristics put forth for Schoolwide Programs. Schoolwide Programs must use effective instructional strategies; Success for All specifies grouping and cooperative learning strategies. Schoolwide programs must assist preschool children in transition; Success for All specifies preschool and kindergarten components. Schoolwide programs must provide opportunities for parent involvement; Success for All schools set up a Family Support Team. Schoolwide programs must provide immediate assistance to students who are having difficulty; Success for All institutes a one-to-one tutoring program and conducts eight-week assessments to identify children who need assistance.

In short, Schoolwide Programs as defined in the new IASA provide a design for schools to use to build effective schoolwide programs. The Success for All program operationalizes this design -- it provides a model program.

The differences between a schoolwide program design and an actual schoolwide program such as Success for All create major problems if we are trying to develop one set of standards for designating both of these entities as exemplary and promising. These problems become even more acute when we look at the variation in other schoolwide and multi-site programs. For example, Levin (1995) describes the Accelerated Schools "model" as follows:

> ... accelerated schools are not a cookie cutter intervention that will determine precisely what the specific curriculum, instructional strategies, and school organization will look like next year.... They represent a process of change ... in which the school develops its own dream, an effective system of problem-solving and decision-making to reach that dream, and the creation of powerful learning situations that build on the strengths of students, parents, and school staff. In this respect, each school will be different.... (p. 3)

Similarly, the Essential Schools principles provide a design that schools can apply, with the principles of the design made deliberately ambiguous so that schools will develop their own unique processes and practices to achieve the principles.

The problematic nature of trying to set a common set of standards for the evaluation of specified versus less-specified programs is made clear by McCollum (1994) and Bodilly (in press). McCollum identified two types of programs: curriculum-based (Reading Recovery, Success for All, and the Academy model) and governance-based (Accelerated Schools, School Development Program, and school-based management projects). She concluded that curriculum-based reform is more predictable and depends more on the skills of individual teachers; whereas governance-based reform has goals that constantly change and depends on sustained leadership and teacher commitment. Bodilly (in press), reporting on RAND's formative assessment of the demonstration phase of the NASDC designs, classified the designs in three categories: Core designs, which emphasize changes in elements associated with the core of schooling; Comprehensive designs, which emphasize more elements, including integrated social services, governance, and organization and staffing; and Systemic designs, which emphasize changes to all elements and the need for collaboration among many partners. These categories are progressively broader and more abstract. In essence, Bodilly finds core designs being implemented in the demonstration phase most successfully, with the comprehensive designs and the one systemic design showing low to moderate progress. It may be that broad, abstract designs require more time to fully implement and thus their effects should be judged later in the process. Conversely, because such designs may be more difficult to implement, they may show early signs of failure.

It is useful to note that NASDC emphasizes that the nine reform efforts that it supports are designs, not programs. Making the distinction between designs and programs could be very useful in any attempt to set realistic standards for designating schoolwide and multi-site programs as promising and exemplary. It is legitimate to require a schoolwide program that consists of specified elements and corresponding materials to show that it can be replicated in other schools; it is not legitimate to require the same of a design that consists of guidelines and principles for schools to build upon to develop their own unique vision, objectives, and effective programs.

Program and/or School Level Designations

It will be important to define the level at which a program can be designated as promising or exemplary. Do we want to try to identify Chapter 1 Schoolwide, Essential Schools, or Accelerated Schools, for example, as a promising or exemplary program per se, based on evaluations from multiple sites? Or should the recognition as promising or exemplary occur only at the school level?

Could Essential Schools be identified as either a promising or exemplary program per se on the basis of evaluations that find good results in some schools, but few results in many others, and the inability to implement the program at all in many others?

OERI may interpret "program" to include broad initiatives, but there will be great difficulty in designating any schoolwide or multiple-site programs as *exemplary* in general, meaning that schools that implement the program can expect to improve student outcomes. Too much of the success of these programs depends upon the implementation, not upon specific program elements, for any of them to be termed exemplary in general. The exception to this may be the Success for All program, due to its specification of elements and its professional development targeted on those elements, but even Success for All may achieve more impact on students in some sites than in others.

Identifying the schoolwide and multiple-site programs as *promising* is less problematic. They are promising, in that all have shown in at least some instances that they can be implemented in schools with resultant impact on student outcomes. It is true, as we noted previously, that a more sophisticated and mature field of promising programs now exists that ever before. It is also true, based on our review of evaluations, that some are more promising than others.

School level designations of promising and exemplary seem more appropriate for the Chapter 1 schoolwide and multi-site programs. A clear scenario is easily seen -- a schoolwide program that shows implementation of the various elements of any one of these designs could be designated as promising. For example, a schoolwide program that offers evidence that it has put in place a number of the eight components of a Title I schoolwide program or the nine components of the Comer School Development Program or the three instructional approaches of the Paideia Program could be considered a promising program. The potential is there, based on research, for improved student outcomes.

To be designated as exemplary, the schoolwide or multi-site program would then need to provide convincing evidence of improved student learning. The problems involved in measuring such achievement are discussed next.

Another issue needs to be considered in designating school-level programs as promising or exemplary -- the level of schooling. Almost all Title I schoolwide and multi-site programs that our evaluations find to be effective for improving student achievement are elementary school programs, with some slight movement into the middle grades. There is a lack of Chapter 1 schoolwide programs for high schools -- notably, a session devoted to the topic at the 1994 National Schoolwide Project Conference in Baltimore was cancelled. Evaluations of the high school program, Essential Schools, reveal the large difficulties of design and implementation that occur at the high school level. The guiding principles of Essential Schools have as much face validity and individual research base as most of the other programs we've reviewed, and the program itself has had extensive funding, offers professional development, and has received state-level support -- yet this program's implementation problems are well documented (Viadero, 1995). In the designation of promising and exemplary programs and in the structuring of expert panels, some attention needs to be paid to the special contexts that are faced by high school programs.

Achievement Measurement Issues

We can expect that, in the end, definitions of promising and exemplary programs and the distrinction between them will depend heavily upon whether a program can show evidence of effectiveness for improving student achievement -- reading, math, science, writing, and so on. This "bottom line" standard, however, is in a state of flux among schoolwide and multiple site programs.

For schoolwide Title I (Chapter 1) programs, current evaluations of effectiveness are based extensively on norm-referenced reading and mathematics assessments. As noted, some district-wide evaluations and numerous school case-study evaluations report normal curve equivalent (NCE) gains that show sufficient evidence of effectiveness to be perhaps considered as exemplary. There is a problem in that the NCE gains are generally compared to those of traditional Chapter 1 schools (those with nonschoolwide programs) and not to other schools; nonetheless, these evaluations do reflect genuine program impact on students.

In the Improving America's Schools Act, however, the measurement of impact for Title I students and for schoolwide programs has changed. Impact will now be measured by student performance on the new measures that each state develops to meet its new state-developed standards. The achievement of Title I students will now be measured more on "performance assessments" -- measures that better reflect students' capabilities for critical thinking and problem solving. At the same time, their performance on these assessments will be compared to the performance of all other students in the state.

These changes obviously have large ramifications for the designation of any Title I schoolwide program as a promising or exemplary program. In essence, the main evidence of effectiveness for improving student achievement provided by current Chapter 1 schoolwide programs (i.e., standardized test scores) may be irrelevant. At the same time, many multiple site programs that have some evidence of effectiveness also cite student performance on standardized test measures, not on performance-based assessments.

The resolution of measurement issues in designating exemplary programs will not be easy. Portfolio assessments, for example, have been shown to be useful in improving instructional processes and student performance in schools, but not as useful in comparing the performance of one school or one district with any other (Koretz et al, 1994). We expect that one of the major issues facing the new joint review panel will be determining effectiveness of individual programs in the face of new school-specific and time-specific measures such as portfolios. Providing evidence of effectiveness will less often be a case of conducting fairly rigorous comparisons based on more easily-understood standardized test measures.

However, the new requirement of applying state high-standards assessments to evaluate student performance in schoolwide programs does show large potential for providing measures of real effectiveness. The state of Maryland offers an example of how statewide performance assessment measures based on high standards can be used to determine the effectiveness of Title I schoolwide and multi-site programs. The MSPAP (Maryland State Performance Assessment Program) is a statewide testing program in which all Maryland schools participate. Testing is carried out once per year in grades 3, 5, and 8 (testing at the high-school level is to be added) in reading, language, math, science, writing, and social studies. The tests are perfomance assessments based on high standards -- students are required to write, to solve problems, and to think critically. The test results are reported at state, district, and individual school levels, and are reported as the percentage of students who achieve at levels of not satisfactory, satisfactory, or excellent. For schoolwide programs and for multi-site programs especially in elementary schools that are schoolwide, the individual school results reported each year provide a strong measure of school performance that can be compared to previous years' results and to the results of other schools throughout the district and state.

Still, the determination of evidence of student achievement that deserves to be called "exemplary" remains an issue. For example, this type of performance evaluation has been conducted over a three-year period in four Maryland schools that have implemented the Roots and Wings program, which is one of the NASDC designs. The assessments show the following results for these schools from 1993-95: In language, in 1993, 17% and 23% of third- and fifth-graders respectively performed at the satisfactory or excellent level; in 1995, 39% and 38% respectively did so. In science, in 1993, 19% and 28% of third- and fifth-graders respectively performed at the satisfactory or excellent level; in 1995, 45% at each grade level did so. Results are similar for the other subjects tested, and these comparative numbers reflect impressive gains in student achievement on performance assessments based on high standards. Gains in comparison to the statewide average and to the district as a whole are also impressive.

It remains problematic, however, that in most cases, the majority of students in all schools perform at a less than satisfactory level on these assessments, and few schools have yet reached the goals set by the state of Maryland on these assessments -- 70% of the school's students achieving at or above satisfactory, and 25% of the school's students achieving at or above excellent. If we use these standards of excellence on difficult performance assessments, such as the achievement of the 70-25 goal, for designating a school's program as exemplary, many schoolwide and multi-site programs that produce large increases in student achievement might nonetheless never be designated as exemplary. Thus a major problem to be confronted as we attempt to set standards for promising and exemplary programs based on new state assessments will be to determine degrees of achievement on the various measures that reflect major improvement, even if that improvement might not reach state goals.

PART III: IMPLICATIONS FOR THE DESIGN OF STANDARDS

In this section, we discuss the implications of our review and discussion of evaluations of schoolwide and multi-site programs for setting standards for the designation of exemplary and promising programs. This discussion involves the definition of an educational program, priorities for programs, the membership and operation of expert panels, and the standards of need and significance, quality, effectiveness, and usefulness to others.

What is an educational program?

Our review of schoolwide and multi-site programs finds two (and maybe three) distinct types of programs that might not be well served by having one set of standards applied to all. The basic distinction is between well-specified programs accompanied by sets of materials (referred to by

Bodilly as "core" programs and by McCollom as "curriculum-based"), and programs that provide assistance and guidance to schools in developing their own procedures, practices, and materials (referred to by Bodilly as "comprehensive designs" and by McCollom as " governance-based"). Actually, the former might be best classified as "programs" and the latter be best classified as "designs." "Programs" can be and should be replicated by schools to the extent possible. "Designs" can be implemented in schools to create specific school programs, but the programs differ from school to school. Programs can be identified as exemplary based on the achievement effects of schools that replicate them. Designs can be identified as promising based on the achievement effects of schools that are able to build effective programs based on the design.

We suggest that OERI consider the need to set standards that recognize the major difference between replicable programs and programs that are not and do not want to be replicable. The use of the term "program" for one and "design" for the other may be one way to do this.

Our review of schoolwide and multi-site programs also raises the question of whether distinctions need to be made in the standards applied to designating elementary and secondary school programs -- especially high school programs -- as promising or exemplary. We're not suggesting that lower standards should be applied to high school programs -- we do suggest that the development and implementation of successful schoolwide or multi-site programs at the high school level seems to face greater obstacles than at the elementary school level, and the greater difficulty involved needs to be recognized.

Priorities for programs

The identification of schoolwide and multi-site programs should be a major and separate priority upon which to focus the review of exemplary and promising programs. Actually, schoolwide alone as a priority might suffice, as the multi-site programs we have examined are themselves schoolwide programs that have been implemented in large numbers of sites.

Subject-matter priorities (i.e., effective math, science programs), governance priorities (i.e., effective collaborative management structures), or other more specific priorities in most cases are all addressed by schoolwide programs. However, a focus on identifying exemplary and promising programs in these specific areas (math and science, for example) would probably exclude most schoolwide programs (even though part of their effects include improved student achievement in math and science).

We suggest that schoolwide programs be a specific priority area in the review of exemplary and promising programs.

Membership and operation of expert panels

Given the major differences between evaluating schoolwide programs and evaluating subject-area or classroom-based programs, the major differences in evaluating various types of schoolwide programs in themselves, and the current national growth in the implementation and dissemination of schoolwide programs, we would suggest *establishing a schoolwide program expert panel*. Consider the possibility of two panels, one for elementary school schoolwide programs, and another for secondary school schoolwide programs.

Panel expertise should include knowledge of alternative assessment, authentic assessment, and performance assessment, as many future evaluations of schoolwides should be conducted using these types of measures.

The Standards for Designating Programs as Exemplary and Promising

Need

Existing schoolwide and multi-site programs are based on clear and demonstrated educational needs for improved student learning in all areas throughout an individual school, especially for students who are considered at-risk -- those students who are achieving least well in our current schools. In essence, schoolwide programs respond to the recognition that a systemic schoolwide initiative is more effective than individual programs related to subject areas. Therefore, standards related to the need for schoolwide and multi-site programs should include:

- the program responds to the need for improved student achievement in multiple subject areas in schools in which previous achievement levels have been low.

Significance

Evaluations of schoolwide and multi-site programs have shown that these programs address a nationally significant educational problem of low achievement schoolwide, especially in schools that have a majority of at-risk students. They develop and combine multiple innovative strategies, practices, and classroom programs that generally have a research base and which serve the diversity of student populations within the schools. Most schoolwide and multi-site programs address student diversity directly, and include components that seek to build upon cultural strengths of students. These programs, in general, also include structures and components that bring together all stakeholders -- parents, community, teachers, and administrators -- in a coordinated effort to improve the school as a whole and student achievement throughout the school. Therefore, standards related to the significance of schoolwide and multi-site programs should include:

- the program has the potential to contribute knowledge to solving the current problem of low student achievement in individual schools.

- the program embodies and/or builds upon school and classroom practices and innovations that have prior evidence of effectiveness.

- the program addresses student diversity through components that seek to build on the strengths of that diversity.

- the program provides structures for systemic schoolwide reform, seeking to involve all stakeholders and providing mechanisms for their involvement.

Quality

Descriptions of the principles, guidelines, and specific components of schoolwide and multi-site programs show that they generally have multiple characteristics that indicate quality. All have a very clear goal -- the improvement of student learning in the whole school. In general, they are derived from a conceptual basis in research or best practice, although some are based on much stronger research evidence than others.

In general, schoolwide and multi-site programs share a conception of organization, curriculum, instruction, and assessment that currently are probably best reflected in the standards of quality provided for Title I schoolwide programs in the Improving America's Schools Act. Pechman and Feister (1996) identify components of this list that, based on their evaluations, seem to reflect high quality, and our review of evaluations supports their selected componets. Thus we suggest the following standards for the quality of programs to be designated as exemplary and promising. Such programs include:

- an agreed-upon vision and goals;

- time and resources for planning and program design;

- a well-defined management and organizational structure;

- a clear focus on academics;

- continuing professional development that includes all stakeholders;

- a commitment to cultural inclusiveness;

- parent and community involvement; and

- appropriate accountability measures.

Effectiveness

The establishment of standards of effectiveness for improved student learning for schoolwide and multi-site programs is a complex issue being made even more complex by the movement of states, districts, and schools away from standardized norm-referenced measures and toward more appropriate performance measures. As noted, the IASA mandates that Title I schoolwide programs will now be evaluated on criteria that reflect state measures of actual student performance in critical thinking and problem solving, as opposed to the ability to respond well on tests of recall of information. At the same time, many multi-site programs have stated a preference for evaluation on these types of measures, even while conducting evaluations that apply standardized norm-referenced measures and criterion-referenced measures that are not performance based.

If criteria of effectiveness for improved student learning are to be based on the new measures that reflect high standards in student performance, multiple problems emerge. First, the current evaluations of many schoolwide and multi-site programs that show effectiveness have been conducted using non-performance based measures. Second, most states are in the beginning stages only of setting their new high standards and developing their new performance measures. Third, the goals that states set for designating satisfactory and excellent school performance on these measures must be reconciled with criteria that designate satisfactory and excellent student progress toward those goals, or few if any schoolwide or multi-site programs are likely to be designated as effective within the next five years or so.

In short, schoolwide and multi-site programs are betwixt and between in terms of how they should be evaluated for effectiveness in improving student learning. It would be unfair to designate criteria for effectiveness based only on evaluations using new performance measures. At the same time, these types of evaluations should be encouraged as the most appropriate (as they are by IASA), and their application to issues of effectiveness should be studied and clarified. As noted, expert panels that have responsibility for examining schoolwide and multi-site programs and recommending action on them need to have special expertise in the area of assessment in general and especially in new methods of performance, portfolio, and authentic assessment measures.

Given the myriad concerns involved in the assessment of whether a schoolwide or multi-site program contributes to improved student learning, it would probably be best to apply a standard for designating effectiveness in improving student learning that remains general enough to include multiple types of assessment, such as the current draft statement:

- The program contributes to improved student learning and can relate such improvement to its activities.

This general type of statement provides an expert panel with the latitude to consider various assessment measures on a case-by-case basis. We would suggest, however, that the use of performance measures be encouraged and rewarded, and that the statement be revised in the future (say five years from now) to say:

- The program contributes to improved student learning as measured by performance, portfolio, or authentic assessments based on high standards, and can relate such improvement to its activities.

Usefulness to others

The draft standards for the designation of exemplary and promising programs, as they are, seem to be appropriate for schoolwide and multi-site programs in specifying that such programs may benefit others beyond the sites in which they were originally developed through "use," "replication," or "adaptation." These standards for usefulness address the variation in how these programs are implemented in additional sites. Cost and staff development provisions also are appropriate.

Exemplary and Promising Status

We would suggest that positive effects on student learning should be the criterion that distinguishes exemplary programs and promising programs, and that the effects on student learning be shown by assessments that include standardized norm-referenced and criterion-referenced tests and performance, authentic, and portfolio assessments.

REFERENCES

Accelerated School Team (n.d.). "Edison Accelerated School." Effective Practices Specialty Option, Chapter 1 Bulletin Board System, PRC Inc., Indianapolis IN.

Adler, M. (1984). *Paideia Proposal: An Educational Manifesto.*

Bodilly, S. (in press). "Lessons learned from RAND's formative assessment of NASDC's phase 2 deomonstration effort." In Stringfield, S., Ross, S., and Smith L. (Eds.), *Bold plans for school restructuring: The New American Schoools Development Corporation models.* Hillsdale NJ: Lawrence Erlbaum.

Boger, J.M. (1989). *Psychosocial and academic effects of the School Development Program.* Unpublished Dissertation, Teachers College, Columbia University, New York.

Cauce, A.M., Comer, J.P., and Schwartz, B.A. (1987). "Long-term effects of a system-oriented school prevention program.' *American Journal of Orthopsychiatry, 57*(1): 127-131.

Christener, C.A. (1987). *Schoolwide projects: The almost revolution (?) six years later.* Paper presented at the AERA Annual Meeting in Washington, DC, 1987.

Comer, J.P. (1988). "Educating poor minority children." *Scientific American, 259*: 42-48.

Committee on Education and Labor. (1990). *Chapter 1 survey of the Hawkins-Stafford School Improvement Amendments.* Washington, DC: US House of Representatives.

Cushman, K. (1991). "Taking stock: How are Essential Schools doing?" *Horace, 8*(1): 1-12.

Davidoff, S.H. and Pierson, E.M. (1991). *A continued look at the promise of schoolwide projects.* Paper presented at the AERA Annual Meeting in Chicago. Philadelphia, PA: The School District of Philadelphia Office of Assessment, Educational Programs Evaluation Unit, Testing Services Unit.

Desmond, C.T. (1992). *A comparison of the assessment of mastery in an outcome-based school and a Coalition of Essential Skills school.* Paper presented AERA Annual Meeting in San Francisco. (ED347655).

Doss, D. and Holley, F. (1982). *A cause for national pause: Title 1 schoolwide projects.* Austin, TX: Austin Independent School District, Office of Research and Evaluation (ED 214 996).

Education Commission of the States (1995). New American Schools: An Introduction. Denver CO: Author.

Hampel, R.L. (in press). "Janice, Pam, Dave, and Steven: the micropolitics of RE:Learning." *Teachers College Record.*

Haynes, N.M. and Comer, J.P. (1990). "The effects of a School Development Program on self-concept." *The Yale Journal of Biology and Medicine, 63*: 275-283.

Haynes, N.M. and Comer, J.P. (1991). *Summary of School Development Program (SDP) effects.*

Haynes, N.M. and Comer, J.P. (1993). "The Yale School Development Program: Process, outcomes, and policy implications." *Urban Education,28*(2): 166-199.

Haynes, N.M., Comer, J.P., and Hamilton-Lee, M. (1985). *Longitudinal assessment of SDP effects*. New Haven: Yale Child Study Center.

Haynes, N.C., Comer, J.P., and Hamilton-Lee, M. (1988a). The effects of parental involvement on student performance. *Educational and Psychological Research, 8*(4), 291-299.

Haynes, N.M., Comer, J.P., and Hamilton-Lee, M. (1988b). "The School Development Program: A model for school improvement." *Journal of Negro Education*, 57(1): 11-21.

Haynes, N.M., Comer, J.P., and Hamilton-Lee, M. (1989). "School climate enhancement through parental involvement." *Journal of School Psychology, 8*(4): 291-299.

Haynes, N.M. and Emmons, C.L. (1990a). *School climate report: Clemente Middle School.* New Haven: Yale University, Yale Child Study Center.

Haynes, N.M. and Emmons, C.L. (1990b). *School climate report: Fair Haven Middle School.* New Haven: Yale University, Yale Child Study Center.

Haynes, N.M. and Emmons, C.L. (1990c). *School climate report: Jackie Robinson Middle School*. New Haven: Yale University, Yale Child Study Center.

Herman, R. and Stringfield, S. (April 1995). "Ten promising programs for educating disadvantaged students." Paper presented at the annual convention of the American Educational Research Association, San Francisco.

Joyner, E.T. (1990). *The Comer model: School improvement for students at risk*. Unpublished dissertation, University of Bridgeport, Bridgeport, CT.

Hoff (October 20, 1995). "Schoolwide projects sweep newly eligible Title I schools." *Education Daily 28*(203), 1, 3.

Kentucky Department of Education (April 1995). *Schoolwide program planning guide 1995-96.* Office of Special Instructional Services, Division of Program Resources.

Koretz, D., et al (1994). *The Vermont portfolio assessment program: Findings and implications.* Washington D. C.: RAND Institute on Education and Training.

Leighton, M. S., Hightower, A.M. and Wrigley, P.G. (1995). *Model strategies in bilingual education: Professional development.* Washington D.C.: U. S. Department of Education.

Le Tendre, M. J. (1996). "Title I schoolwide programs: Improving schools for all children." Journal of Eduction on Students Placed At Risk 1(2), 109-11.

Levin, H.M. (1995). "Accelerated visions." Accelerated Schools 4(2): 2-3.

Madden, N., Slavin, R.E., Karweit, N.L., Dolan, L.J., and Wasik, B.A. (1993). "Success for All: Longitudinal effects of a festructuring program for inner-city elementary schools." *American Educational Research Journal, 30*(1): 123-148.

McCollum, H. (1994). *School reform for youth at risk: Analysis of six change models. Volume I: Summary and analysis.* Washington D.C.: Policy Studies Associates.

Muncey, D.E. (1994). *Individual and schoolwide change in eight Coalition Schools: Findings from a longitudinal ethnography study.* Paper presented at AERA Annual Meeting in New Orleans.

Muncey, D.E. and McQuillan, P.J. (1993). "Preliminary findings from a five-year study of the Coalition of Essential Schools." *Phi Delta Kappan*, -(-): 486-489.

Pechman, E.M. and Fiester, L. (1994). *Implementing schoolwide projects: An idea book for educators.* Washington D.C.: U.S. Department of Education.

Pechman, E.M and Fiester, L. (1996). "Creating good schools for children in poverty through Title I schoolwide programs." Journal of Research on the Education of Students Placed At Risk, 1 (2), 173-94.

Prestine, N.A. (1993). "Feeling the ripples, riding the waves: Making an Essential School." in Murphy, J. and Hallinger, P. (eds.) *Restructuring schooling: Learning from ongoing efforts*. Newbury Park, CA: Corwin Press, Inc.

Prestine, N.A. (1994). *Sorting it out: A tentative analysis of Essential School change efforts in Illinois*. Paper presented at the AERA Annual Meeting in New Orleans.

Prestine, J.A. and Bowen, C. (1993). "Benchmarks of change: Assessing Essential School restructuring efforts." *Educational Evaluation and Policy Analysis, 15*(3): 298-319.

Ross, S.M. and Smith, L.J. (1992). *Final report: 1991-92 Success for All program in Memphis, Tennessee.* Memphis: Memphis State University.

Ross, S.M., Smith, L.J., Casey, J., and Slavin, R.E. (1994). *Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs.*

Shirley, J.R. and Anderson, L.W. (1994). *A study of South Carolina's project RE:Learning.* Paper presented at AERA Annual Meeting in New Orleans.

Sikorski, M.F., Wallace, T., Stariha, W.E., and Rankin, V.E. (1993). "School reform and the curriculum." *New Directions for Program Evaluation, 59.*

Slavin, R.E., Karweit, N.L., and Wasik, B.A. (in press). *Preventing early school failure: Research, policy, and practice.*

Slavin, R. and Madden, N.A. (1994). *Lee Conmigo: Effects of Success for All in bilingual first grades.* Paper presented at AERA Annual Meeting in New Orleans.

Slavin, R.E., Madden, N.A., Dolan, L.J., Wasik, B.A., Ross, S., and Smith, L. (1994). *Success for All: Longitudinal effects of systemic school-by-school reform in seven districts.* Paper presented at AERA Annual Meeting in New Orleans.

Slavin, R.E., Madden, N.A., Dolan, L.J., Wasik, B.A., Ross, S., and Smith, L. (1993). *Whenever and wherever we choose...The replication of Success for All.* Paper presented at AERA Annual Meeting in Atlanta, GA.

Slavin, R.E., Madden, N.A., Karweit, N.L., Dolan, L.J., and Wasik, B.A. (1992). *Success for All: A relentless approach to prevention and early intervention in elementary schools.* Arlington, VA: Educational Research Service.

Slavin, R.E. and Yampolsky, R. (1992). *Success for All: Effects on students with limited English proficiency: A three-year evaluation.* Baltimore: Center for Research on Effective Schooling for Disadvantaged Students.

Stapleford, T.A. (1994). *Locally initiated vs. nationally based school reform: A comparative study of two school reform projects.* Paper presented at AERA Annual Meeting in New Orlean.

Stringfield, S., Millsap, M., Yoder, N., Schaffer, E., Nesselrodt, P., Gamse, B., Brigham, N., Moss, M., Herman, R., and Bedinger, S. (in progress). *Special strategies studies final report.* Washington, D.C.: U.S. Department of Education.

Stringfield, S., Winfield, L., Millsap, M., Puma, M., Gamse, B., and Randall, B. (1994). *Urban and suburban/rural special strategies for educating disadvantaged children: First year report.* Washington, D.C.: U.S. Department of Education.

Turnbull, B.J., Zeldin, S., and Cain, T. (1990). *State administration of the amended Chapter 1 program.* Washington, DC: Policy Study Associates.

U. S. Department of Education (1992). *Compensatory education programs: National Chapter 1 schoolwide projects conference: Final report.* Washington DC: Author.

U. S. Department of Education (1994). *Compensatory education programs: National Chapter 1 schoolwide projects conference: Conference notes.* Washington DC: Author.

U.S. Department of Education. (1993). *Reinventing Chapter 1: The current Chapter 1 program and new directions. Final report of the national assessment of the Chapter 1 program.* Washington, DC: Author.

Viadero, D. (1995). "Mixed record for Coalition Schools is seen." (November 1995). Education Week XV(9), 1,12.

Wallace, T. (1993). "Chicago public schools: Evaluation of the 1987-88 Paideia program." In D.R. Waldrip, W.L. Marks, and N. Estes (eds.) *Magnet school policy studies and svaluations*. Houston, TX: International Research Institute on Educational Choice, 477-515.

Wasik, B. and Slavin, R.E. (1993). "Preventing early reading failure with one-to-one tutoring: A review of five programs." *Reading Research Quarterly, 28*(2):179-200.

Winfield, L.F. and Hawkins, R. (1993). *Longitudinal effects of Chapter 1 schoolwide projects on the achievement of disadvantaged students.* Report No. 46. Baltimore MD: Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University.

Winfield, L.F., Hawkins, R., and Stringfield, S. (1992). *A description of Chapter 1 schoolwide projects and effects on student achievement in six case study schools (Report no. 37).* Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University.

M

CHANGING THE NATIONAL DIFFUSION NETWORK TO ADDRESS WHOLE-SCHOOL REFORMS

by

Susan Bodilly
Thomas Glennan

October 1996

© Copyright 1996 RAND

CHANGING THE NATIONAL DIFFUSION NETWORK TO ADDRESS WHOLE-SCHOOL REFORMS

by

Susan Bodilly and Thomas Glennan

This paper was commissioned by the Department of Education (ED) as part of an effort to solicit ideas for replacing the old National Diffusion Network with a new system in which panels convened by ED would confer status on newly developed projects and programs. This status would somehow let schools or potential adopters know the "success level" of the project or program in producing improved school or student outcomes, thus promoting more genuine and efficient reform. In particular we were asked to address how the expert panel system now under development should approach and judge the whole-school reforms we have studied for the last several years.

We address this issue in a five parts:

- We examine some aspects of the old system to uncover its flaws and some important lessons so as to guide future development into useful tracks. This is important because "government travels in well worn paths." We must understand explicitly why those paths did not get us where we wanted to go.
- We review some of the findings from others' recent work on the need for systemic reform to point out the fallacies of the old system and understand some of the qualities of reform that must be taken into account by the new system.
- We review our own findings on the New American Schools' effort at promoting whole school reform and draw out implications for characteristics of a new structure.
- We discuss some standards for design-based assistance.
- We discuss the applicability of the ED proposed notions of "promising" and "exemplary".

In general we support moving from the old system embedded in a philosophy of consumer protection, uniformity, and guarantees of success by making projects "teacher proof". That approach is contrary to the whole-school reforms we have studied. Our work in assessing different school- and system-level reforms, especially the RAND assessment of the New American Schools'(NAS) efforts at school-level reform, is the base of observation for arguing for a new role for ED in diffusing reforms. Our suggested approach would focus on informing and empowering schools and teachers to make their own decisions about complex reform and to take responsibility for those decisions.

ISSUES RAISED BY THE PAST APPROACH TO DIFFUSION OF REFORMS

The prior approach to diffusion of reform ideas of the Department of Education, known as the National Diffusion Network (NDN), was designed with an important goal in mind: the successful replication of effective innovations by schools and teachers. Because of the focus on replication, the model for evaluation of innovations, or the conferring of status within the network, was based on scientific methods using experts to review empirical evidence concerning an innovation. This evaluation model is particularly appropriate for replication purposes. By focusing on a narrow reform and holding all else equal, one can measure the effect of the innovation. When exactly replicated, its results can be guaranteed.

Behind this program framework was the theme of consumer protection. The standards for approval, embedded in replication and experimental design, were rigorous to ensure that those programs or projects that passed approval and entered into the network would have positive results in most contexts. Schools or teachers accessing the NDN would be provided only with projects that were proven to reduce the risks of adoption by ensuring replicable results.

Embedded in the approach were assumptions as follow.

- Teachers should not or could not be responsible for improvements in their field. Rather, experts would develop innovations that were teacher proof, i.e. that could be replicated without failure. The adopter was protected from failure by the rigorous process of development and evaluation.

- Effectiveness was best measured by increases in student outcomes largely captured in test scores on normalized, multiple choice assessments. Other measures of effectiveness were largely limited to reduced costs.
- Lack of information about effectiveness prevented the further dissemination of sound educational improvements. Guarantees of effectiveness disseminated through an electronic network and other federal contact points would increase the successful adoption of effective practice.

This system garnered few accolades (Katzenmeyer, 1993, Klein, 1993). A total 150 projects, of limited scope, passed the evaluation process. These tended to be narrowly defined interventions--the only ones that could pass the evaluation criteria with guaranteed consumer protection from failed implementation. They might, by and large, fall into the category of first order reform--improvements in efficiency or effectiveness of existing programs and structures (Cuban, 1988). The old approach was eminently suitable for approving narrow prescriptive reforms. If accessed, then the network would have led to the very "Christmas Tree effect" or "programitis" that has been the bane of school improvement efforts.

Importantly, the network did not have a strong access record. When searching for innovations to adopt, few accessed it. The conclusion inevitably drawn was that the approval process needed to be revamped, as did the means for accessing the network.

We argue that some types of reforms challenged this old approach and we argue for the restructuring of the network to be oriented toward consumer information, with the consumer responsible for choice as well as results. A new philosophy is needed, not just revamping, in order to promote school improvement, as opposed to successful replication of innovations. The remainder of this paper discusses our base of observations, some themes from our observations that suggest new roles for ED, and some suggested ways to approach diffusion to be in keeping with reforms now being undertaken in schools.

THEMES FROM SYSTEMIC REFORMS

Over the past several years, we have investigated or been loosely associated with numerous different K-12 reform efforts through RAND assessments including: an assessment of early innovations supporting integrated vocational and academic educational programs in high schools (Bodilly, Ramsey, Stasz, Eden, 1993), a formative assessment of the GE Foundation College Bound Program (Bodilly, Purnell, Hill, 1994), assessments of the use of education technology on student achievement (Glennan, 1995), an assessment of the differences between parochial and public schools (Hill, 1990), and most recently the formative assessment of the New American Schools (NAS) (Bodilly, et al, 1995, and Bodilly, 1996).

The efforts we have studied were usually founded in reaction **against** narrow program improvements (Smith and O'Day, 1990; Tyack, 1990). Most reforms we have assessed explicitly rejected the adoption of narrow improvements to existing programs feeling that the accumulation of these improvements had contributed to "programitis", a "Christmas tree effect", or the fragmenting of the educational package students receive (Hill and Bonan, 1991). In particular, their supporters felt that defining specific instructional or curriculum improvements did not address fundamental issues of incentives and accountability which had kept the program improvement innovations from being adopted, replicated or sustained. Our base of experience reflects the movement in reform away from narrow interventions to support of broader institutional changes designed to support school and student improvements. The following encapsulate some themes from this reform frontier, each of which violates an assumption behind the past ED approach:

- Some reformers now argue for a more systemic approach that emphasizes basic reforms at all levels of the educational system simultaneously (Elmore, 1994; Hill and Bonan, 1991; Murphy and Hallinger, 1993, Smith and O'Day, 1990). The systemic reforms intend new accountability and incentives packages to promote innovation and initiative at all levels.

The approach is geared to comprehensive and coherent school-level reform undertaken by the practitioners through school-improvement processes. By definition, in these circumstances, the outcomes of narrow program adoption cannot be guaranteed as all factors are changing at once.

- Many new reforms are grounded in a strong proactive role of teachers in the development of the interventions through the use of teacher input on site-base management councils, curriculum development committees, and review of assessment practices (Darling-Hammond, 1988; Darling-Hammond, 1992; Fullan, 1994; Liberman, 1991; Rosenholtz, 1989; Kruse and Seashore Louis, 1995). The commitment to teacher responsibility and authority is a reaction to years of teachers rejecting or not implementing expert-designed improvements. Motivating teachers toward increased responsibility and authority is accomplished through changed incentives and training. This stronger role for teachers as professionals and arbiters of classroom practice also implies the inability to guarantee a faithful replication of a prescribed improvement.
- Some argue strongly for the building of teacher networks to promote diffusion of improvements. Because teachers are the ultimate implementors of much of the curriculum, instruction, standards and assessment improvements, they should be the final arbiters of the usefulness of the changes. This reflects findings that, whenever teachers are asked to adopt an innovation, they ask to speak to other teachers who have successfully adopted it. Skepticism about "experts" practical knowledge of classroom conditions runs high (McLaughlin, 1990).
- Many reforms directly attack the current accountability system based in multiple-choice, standardized test scores. Rather they argue for more "authentic assessments" (Newman, Marks, Gamoran, 1996). This is more important as standards promoted by professional disciplinary associations and states have called for content knowledge and application skills in

students. The ability to usefully apply knowledge can only be assessed through complex performance tasks.

- Finally, some argue that to be successful, each school must go through a process of self-assessment and improvement based on factors peculiar to its clientele. Complex organizations such as schools can only change by "learning by doing". Proponents of this view argue for comprehensive school reforms based in school improvement planning paradigms and constructivist learning by teachers and students (National Alliance For Restructuring Education, 1995; Newman, Marks, Gamoran, 1996).

When added together these reform themes reject a replication model based in consumer protection. Careful evaluation of an innovation becomes less likely as the innovation becomes constructed by and embedded in the school itself and the teaching staff takes on greater responsibility and authority. Rather, the more appropriate model for thinking about these types of reforms is one of organizational learning, perhaps aided by strong networks and shared information by the actual practitioners (Fullan, 1994).

THEMES FROM NAS RESEARCH IN PARTICULAR

We now turn to our findings from one study in particular, the RAND assessment of the NAS experience in its developmental phase known as Phase 2 (Bodilly, et al, 1995; Bodilly, 1995) and the first year of experience in its scale-up phase known as Phase 3. NAS, a private, non-profit corporation was founded in 1991 to develop designs for high performing schools. It invested in nine design teams to develop designs and demonstrate them in real schools, prior to scale-up. In the scale-up phase of the NAS effort or Phase 3, NAS and seven teams are working with a handful of chosen jurisdictions. These jurisdictions have committed to using NAS design-based assistance to transform at least 30 percent of their schools in five years. The design teams and NAS act as change agents for the schools and districts. The teams offer both designs and assistance to schools in the transformation process toward high performance. The jurisdictions begin to change their internal regulatory processes, incentives, assessments, and professional development to promote school level excellence. The NAS approach to reform emphasizes: school choice of a design, theme, or focus; development of unique whole-school cultures; and accountability through performance standards. This contrasts to past emphases on top-down mandates, regulations to encourage school uniformity, multiple program adoption, and promotion of students without regard to standards.

The NAS experiment can be thought of as a model of innovation fitted to the above discussed themes of systemic reform. In systemic reform, schools still need help and experts (Bodilly, 1996). Experts in the NAS model do not have the ultimate, replicable solution to a narrow problem, but a comprehensive design based in research and strong assistance through models, materials, and professional development opportunities. Schools then take these designs and assistance and begin transformation. Transformation is not assured, nor is replication. Rather each site guarantees its own results. In this process each site can be helped toward success by experts who provide assistance geared to the needs of the school. The implementation approach is fluid based on the needs of the school, not a replication model.

The following are points we would take from our research, not findings of the RAND evaluation per se, that have application to the question: what is the role of change agent versus the school in this type of reform, and what are the implications for diffusion of the innovation?

Schools Do Not Have Good Search Models For Considering Innovations

Schools do not approach the question of which innovation to adopt in a manner conducive to buying the best product for the money. A careful approach might be manifested by consideration of multiple options after a gap analysis or needs assessment on the part of the school. We, in the NAS Phase 2 study, and others as well (Huberman and Miles, 1984), found that schools adopt innovations for immediate gain and usually based on consideration of only one or two innovations. The search is limited. This is largely the case because districts do not encourage schools to review their entire performance, nor to take the time and effort to review past performance to determine needs.

In the NAS Phase 2 study, we found schools and districts selected designs looking for gain, not recognizing the effort needed to reform or consideration of what might be appropriate reforms. Sites reported that they chose to work with design teams because: they wanted additional resources, especially new technology; the design team provided access to professional development, and/or the legitimacy gained from association with reputable reformers; or, the design (as the school understood it) matched where the school was headed in some vague way. Principals and teachers did not talk of the resources they would have to expend or the work they would have to do to transform their school.

In interviews at more than one site associated with each team, teachers told us that they approved the design after the principal framed the question as, "agree to do this and get discretionary funds or don't agree and get no discretionary funds." When implementation began, the teachers had not committed to the design, only to the receipt of resources and sometimes resented subsequent demands made upon them.

In the final set of Phase 2 interviews with design teams and sites we asked "what worked well to promote the required changes in the school and what did not work well"? When specifically addressing the selection and initial commitment process, general consensus was clear:

- Fully developed designs, documented in easily readable form would be helpful.
- Teams and sites agreed that the team should present the design to the whole school. Both agree that this would help avoid problems of misinterpretation and "selling" by principals or assignment by districts that plagued several teams and schools.
- Teachers should be involved in the decision to accept a design preferable by a formal, anonymous vote by teachers.
- In making a choice, schools should actually enter into a search for a design suitable to their needs.

The Phase 3 effort was quite different. Jurisdictions ran design fairs where schools could learn about different designs and choose among them. In addition, design teams no longer came with money; districts and schools had to pay for the assistance they were to receive. In most cases, design teams required schools to vote to accept or reject a design--the faculty took a formal vote increasing the responsibility of the teacher in the decision.

Our Phase 3 interviews showed more positive results in matching schools to designs, or at least increasing commitment to designs, unrelated to how much money was offered. Some schools still reported the district "forced them" to accept a design. This led to some disgruntlement by school staff and less expressed commitment to the design. But, for the most part schools at least considered alternatvies carefully before choosing a design. We still found that the principal orchestrated a great deal of the effort and held tremendous sway over opinions, even with formal votes. We also found that schools that had more time to make the decision and entered into the process earlier in the school year prior to implementation were more highly committed. These schools said they had the time to: understand the design and learn about it, reallocate some existing resources into new categories important to successful implementation, plan out the professional development activities more thoroughly, and remove some nonessential or conflicting programs from the school.

Implication for New Ed Role: Rather than assume that schools and teachers are good consumers given an expert judgment as to suitable interventions, a role of ED might be to enable schools to be good consumers. This is not accomplished by disseminating literature about approved replicable innovations. Rather this might be accomplished by other types of consumer information and learning practices such as:

- A users guide for strategic planning for schools and development of school improvement plans.
- A users guide for performing a school gap analysis or needs assessment.
- A users guide for questions to ask "experts" or innovation offerors about what their role will be, what resources will be available, etc.

- A teacher network of contacts for those schools and teachers who have undertaken innovative programs or designs and who will provide thoughtful comments to potential adopters.
- A seminar series presented by districts that schools can attend to become better consumers of whole-school approaches.
- The development and dissemination of best practices in putting on "fairs" where schools can learn about new programs or designs and the maintenance of a list of potential offerors.

In short, a role of ED might be to develop good consumers, not act as the arbiter of fine products. This role would be more in keeping with placing the responsibility and authority for change and improvement in the hands and minds of the schools and districts, but enabling them to perform their jobs well. While we have spoken in general terms, the principal of the school remains a strong force. The above suggested guides and aides might be targeted at principals.

Lack of Investment Resources Drives Fragmentation and Programitis

A major obstacle to adopting or adapting the NAS designs is the lack of investment resources available to schools for transformation. All innovation takes discretionary resources and an investment mentality. Our report on the NAS experience in its Phase 2 and 3 is replete with instances where change failed to take place because resources were available, but were in the wrong pot of money to be used. Other researchers point to the devastating effects of this phenomena on reform (Hill and Bonan, 1991).

At this point in time, schools and district budgets remain fragmented into many categories designed to support the existing system of program add-ons. Most schools and districts are not yet prepared to undertake reform, because they are still thinking in terms of, and their budgets reflect, marginal improvements to existing programs. In short, a major barrier to innovation is not good information about effectiveness, but lack of investment funds directed at innovation and improvement.

This was made especially clear in NAS' Phase 3 when districts committed to transforming large numbers of schools. These districts were used to making checks out to vendors, experts, and professional developers for school improvement, but the checks had always been in small amounts. They were buying piece meal services and projects--much like decorating a Christmas tree. And because these projects were paid for out of many different accounts, no one person in the district was accountable for reform. The risk involved was spread over many.

When these districts first saw the price associated with transforming a whole school, they were distressed. It took a critical review of how they had invested their funds in the past before they could understand the importance of what they were buying. While in fact the total amount needed for transformation might be no more than a school currently spends in promoting many programs, the larger amount in a single check carried more political risk for districts, even if they admit it probably had a greater likelihood of pay-off.

Implication for New Ed Role: A new role for ED would be the promotion of an investment mentality at the district- and school-level necessary for innovative improvement, of whatever type, to succeed. Some specifics might include:

- Development of networks of schools and district, sharing information about how to tap existing resources to make investment pools.
- User guides for districts and schools on how to approach budget decentralization and the creation of investment pools.
- Development of a group of budget experts who can audit districts and schools on their current practices and advise them as to strategies to improve their spending.
- Encouragement of consortia of foundations and private, non-profits through use of matching federal funds or other incentives, to develop local "community chests" for educational improvement.

School-Level Implementation and Assistance Information Is Needed

In group interviews at the NAS sites we asked teachers and administrators what interactions and supports they thought they needed to undertake their design tasks. The results of the group interviews were straightforward. Our findings indicate principals and teachers were fairly consistent across all sites in what they felt was needed to promote change, regardless of the particular design. Our findings indicate the following elements would be part of a well received school implementation strategy and echo findings of Gitlin and Margonis (1995), Herman and Stringfield (1995), Muncey (1994), Muncey and McQuillian (1993), Prestine and Bowen (1993), and Rosenholtz (1989).

- An introduction to the design by the team that was compelling or at least clear and that was provided to **all** administrators and staff.
- Relevant training provided to **all** administrators and teachers at the school with behavioral changes or new processes modeled.
- Concrete materials to use in classrooms, committees, or other forums for reform.
- Presence of the design team members to help them and/or presence of a facilitator to aid in their understanding on a day to day basis.
- Teacher teaming to work on design issues or curriculum development.
- Participatory governance to ensure continued support by teachers to the design.
- Teacher time for curriculum development, teacher to teacher interactions, and becoming adept at new behaviors (time for practice at the individual and school level).

In addition, we found that those teams that provided for more of these types of supports or for strong versions of a few of these supports were able to demonstrate more progress in schools toward design goals. The bottom line is schools need assistance to implement NAS designs and probably to implement any complex whole-school improvement.

Phase 3 offered a different perspective on this issue. Districts probably need assistance to implement systemic change as well. As mentioned above, districts needed help understanding what an investment function might be and needed advise on how to dismantle their existing program base to promote school improvement. They also sought: concrete assistance on how to go about decentralization of budget decisions to the school level; bench marking capabilities to determine for themselves the extent of implementation; concrete models for systemic district reform of accountability mechanisms, assessments, etc.

Implication for New Ed Role: When the school level finding is attached to the notion that schools must be responsible for their own improvement and that success cannot be guaranteed or replicated, a new role for ED emerges. Rather than focus on measuring the final school-level effectiveness of an innovation, it could promote better consumers by offering information about what makes for good assistance. This could also be done at the district level. This might take the form of:

- Consumer guides for implementation issues, including questions schools and districts can ask reform providers, and check lists of resources needed.
- Analyses of different implementation issues associated with different types of reform.
- Bench marks for district and school change and comparison case studies as the means to understand the process.

School Improvement Takes Time and Is Difficult to Measure

The NAS experience points to another major barrier to school improvement: expectations for results. Three difficulties were evident. First, NAS, districts, schools and evaluators had short-time horizons for improvement. However, complex school-reform takes considerable time to complete--informal agreement appears to hover around a five year period being needed for measurable change in all areas. However, measurable change can take place in particular areas quickly. Implementation processes are observable well before this time and some limited student achievement changes or comprehensive school climate factors might be evident, but the full design takes time to implement and take effect. As Pringle and Adelman (1995) state:

> "...most mainstream administrators, school boards, and taxpayers underestimate how much time is needed for school faculty members to individually and collectively imagine and examine radically different conditions of schooling; to coordinate efforts to experiment with these new conditions; to reflect on and evaluate these experiments and then institutionalize the most worthwhile, discard the unacceptable, and refine the rest; and to maintain simultaneously the daily functioning of the school".

Second, the reforming NAS schools were caught between old forms of assessments and as yet not fully developed performance-based assessments. The designs were developed in accordance with new standards to ensure more complex levels of learning and performance by students. Yet they were being evaluated based on old forms of assessment that were antithetical to the purposes of the design (Mitchell, 1995).

Third, many districts maintain, due to political pressures, that tests scores are the one and only measure of effectiveness. However, the complex designs were developed to change the schools on many fronts including: greater teacher satisfaction, greater parental involvement, reduced costs, and increased skill levels. Not only did the assessment used by districts ignore many skill areas for students, they ignored many areas of improvement.

The bottom line is that for complex reforms to be judged successful they must be judged fairly.

Implications for New ED Role: Any evaluation or standards adopted or used by the Ed network must use both interim and summative measures geared toward long-term change that are matched to the goals of the reform and are rich enough to capture all aspects of complex reforms. ED can promote and encourage change in the evaluation field by:

- A consumer guide to indicators of change including expectations for change associated with different types of reforms.

NAS Experiences Point to New View of National Diffusion

When taken together these themes from the NAS experience point to additional or new roles for Ed in diffusing reform practices. These new roles are more in keeping with several major reform efforts now underway. They move ED away from a consumer protection, standards driven model to one that promotes consumer responsibility and consumer gains in knowledge. They also require that ED and its expert panels begin to think of ways that it can promote diffusion by connecting to other reform agendas, such as budget decentralization and teacher professionalization through networks.

These suggested changes do not prohibit the existing ED diffusion activities, which would still be helpful to strong schools searching for marginal improvements. Rather, they add a mission that directs attention to those schools searching for the means to transform and requiring assistance to do so.

NEW DIFFUSION STANDARDS FOR DESIGN-BASED, ASSISTANCE TEAMS

The traditional diffusion role of ED has been to provide a stamp of approval for innovations as to their effectiveness and ability to be replicated. In one sense we see that this role could still be useful if this new mission is adopted. But, some major changes would have to take place.

We found that the contribution of NAS to national reform efforts was largely in the concept of the "design-based assistance organization". In this sense NAS' efforts can be seen as a capacity building policy (McDonnell and Grubb, 1995). The notion of a design-based, assistance organization is a set of propositions arrived at inductively, but supported by the experiences of the NAS design teams and sites which indicated the necessary attributes of the organizations that could most effectively aid schools in transformation. We propose that a design-based assistance organization has the following attributes:

- A capable, well staffed design team that understands and supports the tenets of the design,

- A fully developed design that communicates effectively the vision and specific tasks of school reform advocated by the team,
- A proven implementation strategy and capability to provide assistance overtime,
- The existence of demonstration sites to act as further laboratories of reform and to provide hands-on evidence of success.

If these type organizations begin to spread and their use takes hold, then ED might have an additional diffusion role in terms of standards for design-based assistance organizations. ED might consider how it can develop standards for designs and assistance that could be applied to organizations that act as change agents of reform. Passing the standards would not guarantee results, but would guarantee that the organization had attributes and assistance that could lead to results, if matched with good faith efforts on the part of a school or district. Some important considerations follow.

A Capable Design Team That Can Provide Design-Related Assistance

A capable team would include a stable and established leader or set of leaders, an experienced staff dedicated to the ideas of the design and capable of serving the needs of sites, and, when called for, the proven working partnerships described by the design vision. The team would be capable of providing design experts to provide inservices to school and district staff on the uses of the materials, models, and the processes for change. This implies that the design team successfully develops its own people or relies on competent contractors who understand and support the tenets of the design.

A Fully Developed Design That Communicates Effectively the Vision and Specific Tasks of School Reform Advocated by the Team

In every case, the vision of the NAS teams is student-centered and standards-based, but each team uniquely expresses and emphasizes different functions, purposes, and missions of the school. The vision of a design is often found in the focus on **student expectations** with clear and specific goals for education. Those goals are communicated through many means including: definitive standards, strategic planning processes, and curriculum and instruction packages.

The provision of concrete materials expressing the design details help teachers work through the process of change and guided their efforts toward demonstrating design elements. These models should include not just curriculum and instruction, but models of new infrastructure, new scheduling options, new budgeting processes, etc. This holds whether the team relies on the team for development or the site. Either approach should be described in concrete form, with documentation, and specific applications that the average site level staff could understand and use. In the end, it is the completeness of the package of models and materials that communicates the full vision of "whole-school" reform.

An essential part of this package is a self-critical, feedback mechanism to the school which creates the dissonance between desired outcomes and current outcomes needed to motivate change. Every model of organizational change begins with the introduction of a performance gap as essential to motivating people toward change and ends with a continuous feedback loop that promotes continued change. We found that design teams that had successfully inserted a meaningful self-examination into the school organization could show that this component had some likelihood of sustaining the school over time, after a strong design team presence left. This self-critical component manifested itself in different ways:

- Well documented and understood standards and school assessment components.
- Strategic planning processes that collected and reviewed performance information in an on-going basis.
- Visits from teams of colleagues from other schools and outside experts as "critical friends" to review progress made by the site and point out weaknesses and strengths.

- Visits by school staff to schools with characteristics common to their own that were considered by the team to be excellent schools.
- Facilitators associated with design teams, or design team personnel that acted in a similar capacity, to point out weaknesses and strengths and the means for improving according to the design tenets.
- Continued expectations for change communicated by the design team members in their site visits and through the use of published "milestones" for implementation, with associated expectations for progress toward transformation.

A Proven Implementation Strategy

An assistance organization should have a proven, or a set of proven, implementation strategies which would include: (a) a plan for choosing new sites and introducing the design effectively; (b) a capability to define effectively the relationship between the sites and team; (c) the ability to provide information to the sites through the exchange of materials, conferences, inservices, and use of facilitators, etc.; (d) the ability to support the sites in changing the infrastructure of the schools such as new scheduling plans, needed committees, new teacher evaluations, new staffing assignments, (e) the capability to lead sites through a quality control process that would ensure the design was implemented by all teachers and reached all students (implementation milestones); (f) feedback from practitioners that the approach had proved helpful and resulted in improved outcomes; and (g) access to resources to support this work.

The provision of a quality control mechanism that would ensure that all teachers and students were performing well is essential. This is closely related and perhaps inseparable from the self-critical feedback loop. It stands apart only to indicate that teachers needed specific feedback on their products, such as curriculum units or classroom delivery, to determine quality in the sense that those products meet the design expectation. This is different from assessment data that might indicate whether or not the school was improving. At the end of Phase 2 teams were beginning to develop these mechanisms and have continued this development into Phase 3, including:

- Processes for teachers to use standards and assessments to effectively change their curriculum and for schools to review curriculum alignment across grades.
- Processes to review all teacher-developed curriculum for quality.
- Development of common rubrics to be used by all teachers.
- Specific examples of student products and their scoring.
- Changed teacher evaluation structures that focused on the tenets of the design.
- Coaching opportunities for teachers failing to implement the design.
- Processes to remove teachers who were not dedicated to the design principles.

The Existence of Demonstration Sites to Act as Further Laboratories of Reform and to Provide Hands-on Evidence of Success

For Phase 3 the final characteristic of a design-based, assistance organization is access to strong demonstration sites. Design teams and sites have argued this is important and interviews of teachers and principals in Phase 3 have proven them correct for several reasons:

- Credibility--Potential sites must be able to see that the team can deliver on its promises before they will agree to try the design.
- Implementation--New partner sites need access to a pool of administrators and teachers who have already transformed who can give them specific and practical assistance.
- Continual progress--The demonstration sites can continue the progress toward the design vision and eventually prove the full design out with firm evaluation.

Teams and their sites had developed firm notions about this and were actively moving forward with their districts to support work and spread at least portions of the design to other schools or the whole design to other districts.

EXEMPLARY, PROMISING AND RECOGNITION

The traditional assignment of exemplary or promising guaranteeing some level of replication would have little meaning for design-team standards, for effectiveness will be largely in the hands of schools and districts who associate with the teams. Instead, these types of interventions could be changed to "recognition" status meaning the teams are recognized as conferring positive benefits when teachers are supported in their work, investment funds are provided, and problems and solutions sets are matched.

This status could have two probability levels or stages that align with the old notions of promising and exemplary. In keeping with others' suggested changes to standards (Klein, 1993), we understand innovations develop over time as more people test them and improvements to the concepts are made. In the case of the design-based assistance organization, we have seen the organizations improve over time in their approaches to implementation, in the materials and assistance they provide, and in what elements of design they consider fundamental versus marginal.

In an early stage of their development they might be assigned a "promising" status. The criteria for meeting this standard would be that the team: developed a research-based design, showed the ability to affect interim process measures, and showed the capability to assist some portion of schools--say one out of three--attain improved outcomes. The promising status would indicate that the team had successfully shown some probability of an effect, but was considered underdevelopment.

In later stages of development, a team might earn an exemplary status--the criteria being that the organization had shown a higher probability of actual outcomes. These outcomes would not be guaranteed, but would be framed as probabilities of changes in certain categories of outcomes: test scores, attendance rates, drop-out rates, parental support for schools, teacher attendence and satisfaction, etc. The standard at that point might be three out of four schools associated with the team had shown improved outcomes.

SUMMARY

We believe that there is an important role for ED to play in the diffusion of reforms. It is geared toward helping consumers be more aware of choices and be better able to make reasonable choices for their circumstances. This requires a complete rethinking of the past system, moving away from the use of expert panels toward the provision of consumer information for reform efforts. In a sense the role we have defined for ED is intended to help create a market for deliberate school improvement efforts, as opposed to support the existing market for quick fixes.

BIBLIOGRAPHY

Adelman, Nancy and Beverly Pringle, "Education Reform and the Uses of Time", *Phi Delta Kappan*, September 1995, pp. 27-29.

Bodilly, Susan, Susanna Purnell, Paul Hill, *A Formative Assessment of the General Electric Foundation's College Bound Program*, RAND, Santa Monica, 1994, MR-463-GEF.

Bodilly, Susan, Kim Ramsey, Cathy Stasz, Rick Eden, *Integrating Academic and Vocational Education: Lessons from Eight Early Innovators*, RAND, Santa Monica, 1993, R-4265-NCRVE/UCB.

Bodilly, Susan, *Lessons from New American Schools Development Corporation's Demonstration Phase*, RAND, Santa Monica, 1996, MR-729-NASDC.

Bodilly, Susan, Susanna Purnell, Kimberly Ramsey, and Christina Smith, *Designing New American Schools, Baseline Observations on Nine Design Teams*, RAND, Santa Monica, 1995, MR-598-NASDC.

Cuban, Larry, "A Fundamental Puzzle of School Reform", *Phi Delta Kappan*, January 1988, pp. 341-344.

Darling-Hammond, Linda, "Policy and Professionalism", *Building a Professional Culture in Schools*, Ed. Ann Lieberman, Teachers College Press, New York, 1988, pp. 55-77.

Darling-Hammond, Linda, "Perestroika and Professionalism: The Case for Restructuring Teacher Preparation", *Excellence in Teacher Education: Helping Teachers Develop Leraner-Centered Schools*, ed. Robert McClure, National Education Association, Washington D.C., 1992, pp.99-127.

Elmore, Richard, "Incentives for Going to Scale with Effective Practices: Some Implications for Federal Policy and Evaluation Design", *Issues and Strategies in Evaluating Systemic Reform*, Papers Prepared for the U.S. Department of Education Office of the Under Secretary, Planning and Evaluation Service, November 1994.

Fullan Michael, "Turning Systemic Thinking on Its Head", *Issues and Strategies in Evaluating Systemic Reform*, Papers Prepared for the U.S. Department of Education Office of the Under Secretary, Planning and Evaluation Service, November 1994.

Gitlin, Andrew, and Frank Margonis, "The Political Aspect of Reform: Teacher Resistance as Good Sense", *American Journal of Education*, Number 103, August 1995, pp. 377-405.

Tom Glennan, 1994, forthcoming.

Paul Hill, *High Schools With Character*, RAND, Santa Monica, 1990, R-3944.

Hill, Paul, and Josephine Bonan, *Decentralization and Accountability in Public Education*, RAND, Santa Monica, R-4066-MCF/IET, 1991.

Huberman, A. Michael and Matthew Miles, "Rethinking the Quest for School Improvement; Some Findings from the DESSI Study", *Teachers College Record*, Volume 86, number 1, Fall 1984, pp. 34-54.

Herman, Rebecca, and Sam Stringfield, *Ten Promising Programs for Educating Disadvantaged Students: Evidence of Impact*, Center for the Social Organization of Schools, Johns Hopkins University, Presented at the American Association Research Association Meeting, April 19, 1995.

Katzenmeyer, Conrad, "Identifying the Best in Mathematics and Science at the U.S. Department of Education", *Evaluation and Program Planning*, Vol. 16, 1993, pp. 241-244.

Klein, Susan, "Are There Better Ways to Identify and Share the Best Within and Among Federal Education Programs?", *Evaluation and Program Planning*, Vol. 16, 1993, pp. 227-239.

Kruse, Sharon and Karen Seashore Louis, "Teacher Teaming-Opportunities and Dilemmas," *Brief to Principals*, Brief No. 11, Center on Reorganization and Restructuring of Schools, University of Wisconsin-Madison, Spring 1995

Liberman, Ann, et al., *Early Lessons in Restructuring Schools, National Center for Restructuring Education, Schools, and Teaching*, (NCREST), New York, Teachers College, Columbia University, August 1991.

McDonnell, Lorraine and Norton Grubb, *Education and Training for Work: The Policy Instruments and the Institutions*, RAND, Santa Monica, R-4026-NCRVE/UCB, 1995.

McLaughlin, Milbrey, "The RAND Change Agent Study Revisited: Macro Perspectives and Micro Realities," *Educational Researcher*, Volume 19, Number 9, December 1990.

Mitchell, Karen, *Reforming and Conforming, NASDC Principals Talk about the Impact of Accountability Systems on School Reform*, RAND, Santa Monica, 1995, DRU-1242-NASDC.

Muncey, D. "Individual and Schoolwide Change in Eight Coalition Schools: Findings from a Longitudinal Ethnography Study," Paper presented at American Education Research Association Meeting, New Orleans, 1994.

Muncey, D. and P. McQuillian, "Preliminary Findings from a Five Year Study of the Coalition of Essential Schools," *Phi Delta Kappan*, 1993, pp. 486-489.

Murphy, Joseph and Philip Hallinger, *Restructuring Schooling, Learning from Ongoing Efforts*, Corwin Press, Newbury Park CA, 1993.

National Alliance for Restructuring Education, 1995.

Prestine, Nona, and Chuck Bowen, "Benchmarks of Change: Assessing Essential School Restructuring Efforts," *Educational Evaluation and Policy Analysis*, Volume 15, Number 3, Fall 1993.

Newman, Fred, Helen Marks, and Adam Gamoran, "Authentic Pedagogy and Student Performance," *American Journal of Education*, Vol 104, august 1996, pp.281-312.

Rosenholtz, Susan, *Teacher's Workplace, the Social Organization of Schools*, Longman, New York, 1989.

Smith, Marshall, and Jennifer O'Day, "Systemic School Reform," *Politics of Education Association Yearbook*, 1990, pp. 233-267.

Turnbull, Brenda, "Using Governance and Support Systems to Advance School Improvement", *The Elementary School Journal*, Volume 85, Number 3, 1985, pp. 337-351.

Tyack, David, "Restructuring in Historical Perspective: Tinkering toward Utopia," *Teachers College Record*, Volume 92, number 2, Winter 1990, pp. 169-191.

VOLUME II *SECTION IV. Legislation and Regulations*	
Key Parts of the 1994 OERI Authorization relating to Expert Panels	**N**
Draft Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI) -- Designation of Exemplary and Promising Programs" Proposed Rule, June 3, 1996	**O**
Final Federal Register Notice: Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI) -- Designation of Exemplary and Promising Programs, Nov. 1997	**P**

N

TITLE IX--EDUCATIONAL RESEARCH AND IMPROVEMENT

SEC. 901. SHORT TITLE.

This title may be cited as the "Educational Research, Development, Dissemination, and Improvement Act of 1994".
-###-

SEC. 805. APPLICATIONS. PREV UP NEXT SEC. 902. FINDINGS.

This section contains parts of the 1994 Goals 2000 legislation reauthorizing the Department of Education's Office of Educational Research and Improvement. The items checked relate to the System of Expert Panels and OERI roles in identifying and sharing promising and exemplary products, programs and practices.

This section is printed from the Education Department's web site <http://www.ed.gov < with the exact address listed on the bottom of each page.

SEC. 902. FINDINGS.

The Congress finds as follows with respect to improving education in the United States:

(1) A majority of public schools in the United States are failing to prepare students to achieve the National Education Goals. The Federal Government should support an extensive program of educational research, development, dissemination, replication and assistance to identify and support the best responses for the challenges ahead. A significant investment in attaining a deeper understanding of the processes of learning and schooling and developing new ideas holds the best hope of making a substantial difference to the lives of every student in the United States. The Office of Educational Research and Improvement within the Department of Education should be at the center of this campaign in order to coordinate such efforts.

(2) The Federal role in educational research has been closely identified with youths who are socioeconomically disadvantaged, are minorities, belong to a language minority, or have a disability. The Federal commitment to education was sufficient to serve not more than--

(A) in 1993, 1 out of every 6 low-income children in need of preschool education;
(B) in 1990, 3 out of every 5 children in need of remediation;
(C) in 1991, 1 out of every 5 children in need of bilingual education; and
(D) in 1992, 1 out of every 20 youths eligible for assistance under the Job Training Partnership Act.

(3) The failure of the Federal Government to adequately invest in educational research and development has denied the United States a sound foundation of knowledge on which to design school improvements. The educational achievement of minority children is of particular concern because at least half of the public school students in 25 of the largest cities of the United States are minority children, and demographers project that, by the year 2005, almost all urban public school students will be minority children or other children in poverty.

(4) The investment goal of the Federal research, development, and dissemination function should be at least 1 percent of the total amount of funds spent on education.

✓ (5) Nationwide model programs and reliable interventions should be demonstrated and replicated, and for such purposes, programs should be established to conduct research and evaluations, and to disseminate information.

✓ (6) The Office should develop a national dissemination policy that will advance the goal of placing a national treasure chest of research results, models, and materials at the disposal of the education decisionmakers of the United States.

(7) A National Educational Research Policy and Priorities Board should be established to work collaboratively with the Assistant Secretary to forge a national consensus with respect to a long-term agenda for educational research, development, dissemination, and the activities of the Office.

(8) Existing research and development entities should adopt expanded, proactive roles and new institutions should be created to promote knowledge development necessary to accelerate the application of research findings to high priority areas.

(9) Greater use should be made of existing technologies in efforts to improve the educational system of the United States, including efforts to disseminate research findings.

(10) Minority educational researchers are inadequately represented throughout the Department of Education, but particularly in the Office. The Office therefore should assume a leadership position in the recruitment, retention, and promotion of qualified minority educational researchers.

✓ (11) The coordination of the mission of the Office with that of other components of the Department of Education is critical. The Office should improve the coordination of the educational research, development, and dissemination function with those of other Federal agencies.

SEC. 912. OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT.

(a) DECLARATION OF POLICY REGARDING EDUCATIONAL OPPORTUNITY.--

(1) IN GENERAL.--The Congress declares it to be the policy of the United States to provide to every individual an equal opportunity to receive an education of high quality regardless of race, color, religion, sex, age, disability, national origin, or social class. Although the American educational system has pursued this objective, it has not been attained. Inequalities of opportunity to receive high quality education remain pronounced. To achieve the goal of quality education requires the continued pursuit of knowledge about education through research, development, improvement activities, data collection, synthesis, technical assistance, and information dissemination. While the direction of American education remains primarily the responsibility of State and local governments, the Federal Government has a clear responsibility to provide leadership in the conduct and support of scientific inquiry into the educational process.

(2) MISSION OF OFFICE.--

(A) The mission of the Office shall be to provide national leadership in--

(i) expanding fundamental knowledge and understanding of education;
(ii) promoting excellence and equity in education; and the achievement of the National Educational Goals by spurring reform in the school systems of the United States;
✓ (iii) promoting the use and application of research and development to improve practice in the classroom; and
(iv) monitoring the state of education.

(B) The mission of the Office shall be accomplished in collaboration with researchers, teachers, school administrators, parents, students, employers, and policymakers.

(b) PURPOSE AND STRUCTURE OF OFFICE.--

(1) IN GENERAL.--The Secretary, acting through the Office, shall carry out the policies set forth in subsection (a). In carrying out such policies, the Office shall be guided by the Research Priorities Plan developed by the Assistant Secretary working collaboratively with the Board and which has been approved by the Board.

(2) ADMINISTRATIVE STRUCTURE.--The Office shall be administered by the Assistant Secretary and shall include--

(A) the National Educational Research Policy and Priorities Board established by section 921;
(B) the national research institutes established by section 931;
✓ (C) the national education dissemination system established by section 941;
(D) the National Center for Education Statistics; and
(E) such other units as the Secretary deems appropriate to carry out the purposes of the Office.

(3) AUTHORIZED ACTIVITIES.--

(A) OFFICE.--In fulfilling its purposes under this section, the Office is authorized to--

(i) conduct and support education-related research activities, including basic and applied research, development, planning, surveys, assessments, evaluations, investigations, experiments, and demonstrations of national significance;
(ii) disseminate the findings of education research, and provide technical assistance to apply such information to specific problems at school sites;
(iii) collect, analyze, and disseminate data related to education, and to library and information services;
(iv) promote the use of knowledge gained from research and statistical findings in schools,

other educational institutions, and communities;
(v) provide training in education research; and
(vi) promote the coordination of education research and research support within the Federal Government, and otherwise assist and foster such research.

(c) APPOINTMENT OF EMPLOYEES.--

(1) IN GENERAL.--The Assistant Secretary may appoint, for terms not to exceed three years (without regard to the provisions of title 5, United States Code, governing appointment in the competitive service) and may compensate
(without regard to the provisions of chapter 51 and subchapter III of chapter 53 of such title relating to classification and General Schedule pay rates) such scientific or technical employees of the Office as the Assistant Secretary considers necessary to accomplish its functions, provided that--

(A) at least 30 days prior to the appointment of any such employee, public notice is given of the availability of such position and an opportunity is provided for qualified individuals to apply and compete for such position;
(B) the rate of basic pay for such employees does not exceed the maximum rate of basic pay payable for positions at GS--15, as determined in accordance with section 5376 of title 5, United States Code;
(C) the appointment of such employee is necessary to provide the Office with scientific or technical expertise which could not otherwise be obtained by the Office through the competitive service; and
(D) the total number of such employees does not exceed one-fifth of the number of full-time, regular scientific or professional employees of the Office.

(2) REAPPOINTMENT OF EMPLOYEES.--The Assistant Secretary may reappoint employees described in paragraph (1) upon presentation of a clear and convincing justification of need, for one additional term not to exceed 3 years. All such employees shall work on activities of the Office and shall not be reassigned to other duties outside the Office during their term.

(d) AUTHORITY TO PUBLISH.--

(1) IN GENERAL.--The Assistant Secretary is authorized to prepare and publish such information, reports, and documents as may be of value in carrying out the purposes of this title without further clearance or approval by the Secretary or any other office of the Department of Education.
(2) QUALITY ASSURANCE.--In carrying out such authority, the Assistant Secretary shall--

(A) establish such procedures as may be necessary to assure that all reports and publications issued by the Office are of the highest quality; and
(B) provide other offices of the Department of Education with an opportunity to comment upon any report or publication prior to its publication when its contents relate to matters for which such office has responsibility.

(e) BIENNIAL REPORT ON ACTIVITIES OF OFFICE.--The Assistant Secretary shall transmit to the President and the Congress by not later than December 30 of every other year a report which shall consist of--

(1) a description of the activities carried out by and through each research institute during the fiscal years for which such report is prepared and any recommendations and comments regarding such activities as the Assistant Secretary considers appropriate;
(2) a description of the activities carried out by and through the national education dissemination system established by section 941 during the fiscal years for which such report is prepared and any recommendations and comments regarding such activities as the Assistant Secretary

considers appropriate;
(3) such written comments and recommendations as may be submitted by the Board concerning the activities carried out by and through each of the institutes and the national education dissemination system during the fiscal years for which such report is prepared;
(4) a description of the coordination activities undertaken pursuant to subsection
(g) during the fiscal years for which such report is prepared;
(5) recommendations for legislative and administrative changes necessary to improve the coordination of all educational research, development, and dissemination activities carried out within the Federal Government; and
(6) such additional comments, recommendations, and materials as the Assistant Secretary considers appropriate.

(f) RESEARCH PRIORITIES PLAN.--

(1) IN GENERAL.--Working collaboratively with the Board, the Assistant Secretary shall--

(A) survey and assess the state of knowledge in education research, development and dissemination to identify disciplines and areas of inquiry in which the state of knowledge is insufficient and which warrant further investigation, taking into account the views of both education researchers and practicing educators;
(B) consult with the National Education Goals Panel and other authorities on education to identify national priorities for the improvement of education;
(C) actively solicit recommendations from education researchers, teachers, school administrators, cultural leaders, parents, and others throughout the United States through such means as periodic regional forums;
(D) provide recommendations for the development, maintenance, and assurance of a strong infrastructure for education, research, and development in the United States; and
(E) on the basis of such recommendations, develop a research priorities program which shall recommend priorities for the investment of the resources of the Office over the next 5-, 10-, and 15-year periods, including as priorities those areas of inquiry in which further research, development and dissemination--

(i) is necessary to attain the National Education Goals;
(ii) promises to yield the greatest practical benefits to teachers and other educators in terms of improving education; and
(iii) will not be undertaken in sufficient scope or intensity by the other Federal and non-Federal entities engaged in education research and development.

(2) CONTENTS OF PLAN.--

(A) The research and priorities plan described in paragraph (1) shall, at a minimum--

(i) set forth specific objectives which can be expected to be achieved as a result of a Federal investment in the priorities set forth in the plan;
(ii) include recommendations with respect to research and development on cross-cutting issues which should be carried out jointly by 2 or more of the research institutes; and
(iii) include an evaluative summary of the educational research and development activities undertaken by the Federal Government during the preceding 2 fiscal years, which shall describe--

(I) what has been learned as a result of such activities;
(II) how such new knowledge or understanding extends or otherwise relates to what had been previously known or understood;
(III) the implications of such new knowledge or understanding for educational practice and school reform; and
(IV) any development, reform, and other assistance activities which have utilized such

knowledge or understanding and the effects of such efforts.

(B) REPORT.--

(i) Not later than 6 months after the first meeting of the Board and by October 1 of every second year thereafter, the Assistant Secretary shall publish a report specifying the proposed research priorities of the Office and allow a 60-day period beginning on the date of the publication of the report for public comment and suggestions.
(ii) Not later than 90 days after the expiration of the 60-day period referred to in clause (i), the Assistant Secretary shall submit to the Board a report specifying the proposed research priorities of the Office and any public comment and suggestions obtained under such subparagraph for the Board's review and approval.

(g) COORDINATION.--With the advice and assistance of the Board, the Assistant Secretary shall work cooperatively with the Secretary and the other Assistant Secretaries of the Department of Education to establish and maintain an ongoing program of activities designed to improve the coordination of education research, development, and dissemination and activities within such Department and within the Federal Government to--

(1) minimize duplication in education research, development, and dissemination carried out by the Federal Government;
(2) maximize the value of the total Federal investment in education research, development, and dissemination; and
(3) enable entities engaged in education research, development, and dissemination within the Federal Government to interact effectively as partners and take full advantage of the diverse resources and proficiencies which each entity has available.

(h) ACTIVITIES REQUIRED WITH RESPECT TO COORDINATION.--In carrying out such program of coordination, the Assistant Secretary shall compile (and thereafter regularly maintain) and make available a comprehensive inventory of education research, development, dissemination activities, and expenditures being carried out by the Federal Government.
✓ (i) STANDARDS FOR THE CONDUCT AND EVALUATION OF RESEARCH.--

(1) IN GENERAL.--In consultation with the Board, the Assistant Secretary shall develop such standards as may be necessary to govern the conduct and evaluation of all research, development, and dissemination activities carried out by the Office to assure that such activities meet the highest standards of professional excellence. In developing such standards, the Assistant Secretary shall review the procedures utilized by the National Institutes of Health, the National Science Foundation, and other Federal departments or agencies engaged in research and development and shall also actively solicit recommendations from research organizations and members of the general public.
(2) CONTENTS OF STANDARDS.--Such standards shall at a minimum--

(A) require that a process of open competition be used in awarding or entering into all grants, contracts, and cooperative agreements under this title;
(B) require that a system of peer review be utilized by the Office--

(i) for reviewing and evaluating all applications for grants and cooperative agreements and bids for those contracts which exceed $100,000;
(ii) for evaluating and assessing the performance of all recipients of grants from and cooperative agreements and contracts with the Office; and
✓ (iii) for reviewing and designating exemplary and promising programs in accordance with section 941(d);

(C) describe the general procedures which shall be used by each peer review panel in its operations;

(D)

(i) describe the procedures which shall be utilized in evaluating applications for grants, proposed cooperative agreements, and contract bids; and
(ii) specify the criteria and factors which shall be considered in making such evaluations;

√ (E) describe the procedures which shall be utilized in reviewing educational programs which have been identified by or submitted to the Secretary for evaluation in accordance with section 941
(d); and
(F) require that the performance of all recipients of grants from and contracts and cooperative agreements with the Office shall be periodically evaluated, both during and at the conclusion of their receipt of assistance.

(3) PUBLICATION AND PROMULGATION OF STANDARDS.--

(A) The Assistant Secretary shall publish proposed standards--

(i) which meet the requirements of subparagraphs (A), (B), (C), and (D) of paragraph (2) not later than 1 year after the date of the enactment of this title;
(ii) which meet the requirements of paragraph (2)(E) not later than 2 years after such date; and
(iii) which meet the requirements of subparagraph (F) of paragraph (2) not later than 3 years after such date.

(B) Following the publication of such proposed standards, the Assistant Secretary shall solicit comments from interested members of the public with respect to such proposed standards for a period of not more than 120 days. After giving due consideration to any comments which may have been received, the Assistant Secretary shall transmit such standards to the Board for its review and approval.
(C) Upon the approval of the Board, the Assistant Secretary shall transmit final standards to the Secretary which meet the requirements of the particular subparagraphs of paragraph(2) for which such standards were developed. Such standards shall be binding upon all activities carried out with funds appropriated pursuant to subsection(m).

(j) ADDITIONAL RESPONSIBILITIES OF THE ASSISTANT SECRETARY.--In carrying out the activities and programs of the Office, the Assistant Secretary--

(1) shall be guided by the Research Priorities Plan developed by the Assistant Secretary working collaboratively with the Board and which has been approved by the Board;
(2) shall ensure that there is broad and regular public and professional involvement from the educational field in the planning and carrying out of the Office's activities, including establishing teacher advisory boards for any program office, program or project of the Office as the Assistant Secretary deems necessary and involving Indian and Alaska Native researchers and educators in activities that relate to the education of Indian and Alaska Native people;
(3) shall ensure that the selection of research topics and the administration of the program are free from undue partisan political influence;
(4) shall ensure that all statistics and other data collected and reported by the Office shall be collected, cross-tabulated, analyzed, and reported by sex within race or ethnicity and socioeconomic status whenever feasible (and when such data collection or analysis is not feasible, ensure that the relevant report or document includes an explanation as to why such data collection or analysis is not feasible);
(5) is authorized to administer funds to support a single project when more than 1 Federal agency uses funds to support such project, and the Assistant Secretary may act for all such agencies in administering such funds; and
(6) is authorized to offer information and technical assistance to State and local educational

agencies, school boards, and schools, including schools funded by the Bureau, to ensure that no student is--

(A) denied access to the same rigorous, challenging curriculum that such student's peers are offered; or
(B) grouped or otherwise labeled in such a way that may impede such student's achievement.

(k) INDEPENDENT EVALUATIONS.--The Secretary shall enter into one or more contracts for the conduct of an independent evaluation of the effectiveness of the implementation of the provisions of this title. Such evaluations shall be transmitted to the Congress, the President, and the Assistant Secretary not later than 54 months after the date of the enactment of this title.
(l) DEFINITIONS.--For purposes of this title, the following definitions apply:

(1) ASSISTANT SECRETARY.--The term "Assistant Secretary" means the Assistant Secretary for Educational Research and Improvement established by section 202 of the Department of Education Organization Act.
(2) AT-RISK STUDENT.--The term "at-risk student" means a student who, because of limited English proficiency, poverty, race, geographic location, or economic disadvantage, faces a greater risk of low educational achievement or reduced academic expectations.
(3) BOARD.--The term "Board" means the National Educational Research Policy and Priorities Board.
(4) DEVELOPMENT.--The term "development"--

(A) means the systematic use, adaptation, and transformation of knowledge and understanding gained from research to create alternatives, policies, products, methods, practices, or materials which can contribute to the improvement of educational practice; and
(B) includes the design and development of prototypes and the testing of such prototypes for the purposes of establishing their feasibility, reliability, and cost-effectiveness.

✓ (5) DISSEMINATION.--The term "dissemination" means the communication and transfer, through the provision of technical assistance and other means, of the results of research and proven practice in forms that are understandable, easily accessible and usable or adaptable for use in the improvement of educational practice by teachers, administrators, librarians, other practitioners, researchers, policymakers, and the public.
(6) EDUCATIONAL RESEARCH.--The term "educational research" includes basic and applied research, inquiry with the purpose of applying tested knowledge gained to specific educational settings and problems, development, planning, surveys, assessments, evaluations, investigations, experiments, and demonstrations in the field of education and other fields relating to education.
(7) FIELD-INITIATED RESEARCH.--The term "field-initiated research" means education research in which topics and methods of study are generated by investigators, including teachers and other practitioners, not by the source of funding.
✓ (8) NATIONAL EDUCATION DISSEMINATION SYSTEM.--The term "national education dissemination system" means the activities carried out by the Office of Reform Assistance and Dissemination established by section 941.
(9) OFFICE.--The term "Office", unless otherwise specified, means the Office of Educational Research and Improvement established in section 209 of the Department of Education Organization Act.
(10) NATIONAL RESEARCH INSTITUTE.--The term "national research institute" means an institute established in section 931.
✓ (11) TECHNICAL ASSISTANCE.--The term "technical assistance" means assistance in identifying, selecting, or designing solutions based on research to address educational problems, planning, and design that leads to adapting research knowledge to school practice, training to implement such solutions, and other assistance necessary to encourage adoption or application of research.
(12) UNITED STATES; STATE.--The terms "United States" and "State" means each of the 50

PART D--NATIONAL EDUCATION DISSEMINATION SYSTEM

SEC. 941. ESTABLISHMENT WITHIN OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT.

(a) IN GENERAL.--

(1) FINDINGS.--The Congress finds as follows:

(A) In order to improve the American educational system for all students, achieve the National Education Goals, and provide for greater educational equity, policymakers, administrators, teachers, and parents must have ready access to the best information and methods available as a result of educational research and development.
(B) The Office of Educational Research and Improvement should have as one of its primary purposes the dissemination of such information and methods in order to assist the national education reform effort.
(C) All current resources within the Office, the Department of Education, and other agencies that can help accomplish the purposes described in subparagraph (B) should be coordinated by the Assistant Secretary, to the extent practicable, so as to form a systematic process to accomplish such purposes.
(D) Education research has the capacity to improve teaching and learning in our Nation's schools, however, teachers need training in the skills necessary to translate research into practice and to allow teachers to become knowledgeable practitioners and leaders in educational improvement.
(E) Adequate linkages between research and development providers and practitioners are essential to ensuring that research on effective practice is useful, disseminated to and supported with technical assistance for all educators, and that all educators are partners in the research and development process.

(2) PURPOSE.--The purpose of this section is to--

✓ (A) create a national system of dissemination, development, and educational improvement in order to create, adapt, identify, validate, and disseminate to educators, parents, and policymakers those educational programs that have potential or have been shown to improve educational opportunities for all students; and
(B) empower and increase the capacity of teachers to participate in the research and development process.

(3) DEFINITION OF EDUCATIONAL PROGRAM.--For the purposes of this section, the term "educational program" includes educational policies, research findings, practices, and products.

(b) ESTABLISHMENT OF OFFICE.--

(1) IN GENERAL.--There is established within the Office an Office of Reform Assistance and Dissemination (hereafter in this section referred to as the "Dissemination Office") through which the Secretary shall carry out all functions and activities described in this section. Such office shall be headed by a Director who shall be appointed by the Assistant Secretary and have demonstrated expertise and experience in dissemination, including promoting the effective use of research in the classroom.
(2) CERTAIN DUTIES.--The Dissemination Office shall--

(A) disseminate relevant and useful research, information, products, and publications developed through or supported by the Department of Education to schools, educators, parents, and policymakers throughout the United States;

(B) operate a depository for all Department of Education publications and products and make available for reproduction such publications and products;

(C) provide technical and financial assistance to individuals and organizations in the process of developing promising educational programs but who might not, without such assistance, be able to complete necessary development and assessment activities;

(D) coordinate the dissemination efforts of the Office, the regional educational laboratories, the research institutes, the National Diffusion Network, and the Educational Resources Information Center Clearinghouses;

(E) provide training and technical assistance regarding the implementation and adoption of exemplary and promising programs by interested entities;

✓ (F) carry out a program of research on models for successful knowledge dissemination, and utilization, and strategies for reaching education policymakers, practitioners, and others interested in education;

✓ (G) develop the capacity to connect schools and teachers seeking information with the relevant regional educational laboratories assisted under subsection (h), the National Diffusion Network, the Institutes assisted under this section, and the Educational Resources Information Center Clearinghouses; and

(H) provide a biennial report to the Secretary regarding the types of information, products, and services that teachers, schools, and school districts have requested and have determined to be most useful, and describe future plans to adapt Department of Education products and services to address the needs of the users of such information, products, and services.

(3) ADDITIONAL DUTIES.--The Dissemination Office shall carry out a process for the identification of educational programs that work, dissemination through electronic networking and new technologies and the functions and activities performed by the following:

(A) The Educational Resources Information Center Clearinghouses.
(B) The regional educational laboratories.
(C) The Teacher Research Dissemination Demonstration Program.
(D) The Goals 2000 Community Partnerships Program.
(E) The existing National Diffusion Network and its Developer-Demonstrator and State Facilitator projects.
(F) Such other programs, activities, or entities the Secretary determines are consistent with purposes for which the Dissemination Office is established.

(c) IDENTIFICATION OF PROGRAMS.--The Assistant Secretary shall coordinate a process through which successful educational programs are actively sought out for possible dissemination through the national educational dissemination system. Such process shall, at a minimum, have the capability to--

✓ (1) work closely with the Institutes, research and development centers, regional educational laboratories, the National Diffusion Network and its Developer-Demonstrator and State Facilitator projects, learning grant institutions established under the Goals 2000 Community Partnerships Program, Department of Education-supported technical assistance providers, and other entities to identify successful educational programs at the regional, State, local, or classroom level;

✓ (2) review successful educational programs supported by the Department of Education through all of its programs;

✓ (3) through cooperative agreements, review for possible inclusion in the system educational programs administered by the Departments of Health and Human Services (particularly the Head Start program), Labor, and Defense, the National Science Foundation, the Department of the Interior (particularly the Office of Indian Education Programs), and any other appropriate Federal agency; and

✓ (4) provide for an active outreach effort to identify successful educational programs through cooperative arrangements with State and local education agencies, teachers and teacher organizations, curriculum associations, foundations, private schools, institutions of higher education, and other entities that could enhance the ability of the Secretary to identify programs for possible inclusion in the dissemination system.

✓ (d) DESIGNATION OF EXEMPLARY AND PROMISING PROGRAMS.--

(1) IN GENERAL.--The Assistant Secretary, in consultation with the Board, shall establish 1 or more panels of appropriately qualified experts and practitioners to--

(A) evaluate educational programs that have been identified by the Secretary under subsection (c) or that have been submitted to the Secretary for such evaluation by some other individual or organization; and
(B) recommend to the Secretary programs that should be designated as exemplary or promising educational programs.

(2) CONSIDERATIONS IN MAKING RECOMMENDATIONS.--In determining whether an educational program should receive a recommendation under paragraph (1), a panel established under such paragraph shall consider--

(A) whether, based on empirical data, which may include test results, the program is effective and should be designated as exemplary and disseminated through the national dissemination system; or
(B) whether there is sufficient evidence to lead a panel of experts and practitioners to believe that the program shows promise for improving student achievement and should be designated as promising and disseminated through the national dissemination system while the program continues to be evaluated.

(3) REQUIREMENT REGARDING APPROVAL OF PROGRAMS.--In seeking out programs for approval under paragraph (2), the Assistant Secretary shall seek programs that may be implemented at the State, local, and classroom level.
(4) REQUIREMENTS REGARDING PANELS.--

(A) A panel shall not eliminate a program from consideration under this subsection based solely on the fact that the program does not have one specific type of supporting data, such as test scores.
(B) The Assistant Secretary may not designate a program as exemplary or promising unless a panel established under paragraph (1) has recommended that the program be so designated.
(C) The Secretary shall establish such panels under paragraph (1) as may be necessary to ensure that each program identified or submitted for evaluation is evaluated.
(D) Not less than 2/3 of the membership of a panel established under paragraph (1) shall consist of individuals who are not officers or employees of the United States. Members of panels under paragraph (1) who are not employees of the United States shall receive compensation for each day they are engaged in carrying out the duties of the panel as well as compensation for their expenses.

✓ (e) DISSEMINATION OF EXEMPLARY AND PROMISING PROGRAMS.--In order to ensure that programs identified as exemplary or promising are available for adoption by the greatest number of teachers, schools, local and State education agencies, and Bureau-funded schools, the Assistant Secretary shall utilize the capabilities of--

(1) the Educational Resources Information Center Clearinghouses;
(2) electronic networking;
(3) the regional educational laboratories;
(4) the National Diffusion Network;
(5) entities established under the Goals 2000 Community Partnerships Program;

(6) department-supported technical assistance providers;
(7) the National Library of Education; and
(8) other public and private nonprofit entities, including existing education associations and networks, that have the capability to assist educators in adopting exemplary and promising programs.

(f) EDUCATIONAL RESOURCES INFORMATION CENTER CLEARINGHOUSES.--

(1) IN GENERAL.--The Assistant Secretary shall establish a system of 16 clearinghouses having, at a minimum, the functions and scope of work as the clearinghouses had on the day preceding the date of the enactment of this title. The Assistant Secretary shall establish for the clearinghouses a policy for the abstraction from, and inclusion in, the Educational Resources Information Center Clearinghouses system for books, periodicals, reports, and other materials related to education.
(2) ADDITIONAL FUNCTIONS.--In addition to those functions carried out by the clearinghouses on the day preceding the date of the enactment of this title, such clearinghouses may--

(A) periodically produce interpretive summaries, digests, and syntheses of the results and findings of education-related research and development; and
(B) contain and make available to users information concerning those programs designated as exemplary and promising under subsection (d).

(3) COORDINATION OF ACTIVITIES.--The Assistant Secretary shall assure that the functions and activities of such clearinghouses are coordinated with the activities of the Institutes, the regional educational laboratories, learning grant institutions, other clearinghouses supported by the Department of Education, the National Diffusion Network, and other appropriate entities within the Office and such Department.
(4) SPECIAL RESPONSIBILITIES of the secretary.--To assure that the information provided through such clearinghouses is fully comprehensive, the Secretary shall--

(A) require that all reports, studies, and other resources produced directly or by grant or contract with the Department of Education are made available to clearinghouses;
(B) establish cooperative agreements with the Departments of Defense, Health and Human Services, Interior, and other Federal departments and agencies to assure that all education-related reports, studies, and other resources produced directly or by grant from or contract with the Federal Government are made available to such clearinghouses; and
(C) devise an effective system for maximizing the identification, synthesis, and dissemination of information related to the needs of Indian and Alaska Native children.

(5) COPYRIGHT PROHIBITED.--

(A) No clearinghouse or other entity receiving assistance under this subsection may copyright or otherwise charge a royalty or other fee that--

(i) is for the use or redissemination of any database, index, abstract, report, or other information produced with assistance under this subsection; and
(ii) exceeds the incremental cost of disseminating such information.

(B) For purposes of subparagraph (A), the incremental cost of dissemination does not include any portion of the cost of collecting, organizing, or processing the information which is disseminated.

(g) DISSEMINATION THROUGH NEW TECHNOLOGIES.--

(1) IN GENERAL.--The Assistant Secretary is authorized to award grants or contracts in

accordance with this subsection to support the development of materials, programs, and resources which utilize new technologies and techniques to synthesize and disseminate research and development findings and other information which can be used to support educational improvement.

(2) Electronic networking.--

(A) ELECTRONIC NETWORK.--The Assistant Secretary, acting through the Office of Reform Assistance and Dissemination, shall establish and maintain an electronic network which shall, at a minimum, link--

(i) each office of the Department of Education;
(ii) the Institutes established by section 931;
(iii) the National Center for Education Statistics;
(iv) the National Library of Education; and
(v) entities engaged in research, development, dissemination, and technical assistance under grant from, or contract, or cooperative agreement with, the Department of Education.

(B) CERTAIN REQUIREMENTS FOR NETWORK.--The network described in subparagraph (A) shall--

(i) to the extent feasible, build upon existing national, regional, and State electronic networks and support video, telecomputing, and interactive communications;
(ii) at a minimum, have the capability to support electronic mail and file transfer services;
(iii) be linked to and accessible to other users, including State and local education agencies, institutions of higher education, museums, libraries, and others through the Internet and the National Research and Education Network; and
(iv) be provided at no cost (excluding the costs of necessary hardware) to the contractors and grantees described in clause
(v) of subparagraph (A) and to educational institutions accessing such network through the Internet and the National Research and Education Network.

(C) INFORMATION RESOURCES.--The Assistant Secretary, acting through the Office of Reform Assistance and Dissemination, may make available through the network described in subparagraph (A)--

(i) information about grant and contract assistance available through the Department of Education;
(ii) an annotated directory of current research and development activities and projects being undertaken with the assistance of the Department of Education;
(iii) information about publications published by the Department of Education and, to the extent feasible, the full text of such publications;
(iv) statistics and data published by the National Center for Education Statistics;
(v) syntheses of research and development findings;
(vi) a directory of other education-related electronic networks and databases, including information about the means by which such networks and databases may be accessed;
(vii) a descriptive listing of materials and courses of instruction provided by telecommunications partnerships assisted under the Star Schools program;
(viii) resources developed by the Educational Resources Information Center Clearinghouses;
(ix) education-related software (including video) which is in the public domain;
(x) a listing of instructional materials available through telecommunications to local education agencies through the Public Broadcasting Service and State educational television networks; and
(xi) such other information and resources the Assistant Secretary considers useful and appropriate.

(D) EVALUATIONS REGARDING OTHER FUNCTIONS OF NETWORK.--The Assistant Secretary shall also undertake projects to test and evaluate the feasibility of using the network described in subparagraph (A) for--

(i) the submission of applications for assistance to the Department of Education; and
(ii) the collection of data and other statistics through the National Center for Education Statistics.

(E) TRAINING AND TECHNICAL ASSISTANCE.--The Assistant Secretary, acting through the Office of Reform Assistance and Dissemination, shall--

(i) provide such training and technical assistance as may be necessary to enable the contractors and grantees described in clause (v) of subparagraph (A) to participate in the electronic network described in such subparagraph; and
(ii) work with the National Science Foundation to provide, upon request, assistance to State and local educational agencies, the Department of the Interior's Office of Indian Education Programs, tribal departments of education, State library agencies, libraries, museums, and other educational institutions in obtaining access to the Internet and the National Research and Education Network.

(h) REGIONAL EDUCATIONAL LABORATORIES FOR RESEARCH, DEVELOPMENT, DISSEMINATION, AND TECHNICAL ASSISTANCE.--

(1) REGIONAL EDUCATIONAL LABORATORIES.--The Assistant Secretary shall enter into contracts with public or private nonprofit entities to establish a networked system of not less than 10 and not more than 12 regional educational laboratories which serve the needs of each region of the United States in accordance with the provisions of this subsection. The amount of assistance allocated to each laboratory by the Assistant Secretary shall reflect the number of local educational agencies and the number of school-age children within the region served by such laboratory, as well as the cost of providing services within the geographic area encompassed by the region.
(2) REGIONS.--The regions served by the regional educational laboratories shall be the 10 geographic regions in existence on the day preceding the date of the enactment of this title, except that in fiscal year 1996, the Assistant Secretary may support not more than 2 additional regional educational laboratories serving regions not in existence on the day preceding the date of enactment of this Act, provided that--

(A) the amount appropriated for the regional educational laboratories in fiscal year 1996 exceeds the amount appropriated for the regional educational laboratories in fiscal year 1995 by not less than $2,000,000;
(B) each such additional regional laboratory shall be supported by not less than $2,000,000 annually;
(C) the creation of any such additional laboratory region is announced at the time of the announcement of the competition for contracts for all regional educational laboratories;
(D) the creation of a regional educational laboratory that involves the combination or subdivision of a region or regions in existence on the day preceding the date of enactment of this Act in which States in 1 such region are combined with States in another such region does not result in any region in existence on such date permanently becoming part of a larger region, nor result in any such region permanently subsuming another region, nor creates within the continental United States a region that is smaller than 4 contiguous States, nor partitions a region in existence on the day preceding the date of the enactment of this Act to include less than 4 contiguous States included in the region on the day preceding the date of enactment of this Act;
(E) the Assistant Secretary has published a notice in the Federal Register inviting the public, for a period of not less than 60 days, to make recommendations with respect to the creation of

1 or 2 additional regional educational laboratories;
(F) the Assistant Secretary has solicited and received letters of support for the creation of any new region from the Chief State School Officers and State boards of education in each of the contiguous States that would be included in such new region.

(3) DUTIES.--Each regional educational laboratory receiving assistance under this section shall promote the implementation of broad-based systemic school improvement strategies and shall have as such laboratory's central mission and primary function to--

(A) develop and disseminate educational research products and processes to schools, teachers, local educational agencies, State educational agencies, librarians, and schools funded by the Bureau, as appropriate, and through such development and dissemination, and provide technical assistance, to help all students meet standards;
(B) develop a plan for identifying and serving the needs of the region by conducting a continuing survey of the educational needs, strengths, and weaknesses within the region, including a process of open hearings to solicit the views of schools, teachers, administrators, parents, local educational agencies, librarians, and State educational agencies within the region;
(C) provide technical assistance to State and local educational agencies, school boards, schools funded by the Bureau, as appropriate, State boards of education, schools, and librarians;
(D) facilitate school restructuring at the individual school level, including technical assistance for adapting model demonstration grant programs to each school;
(E) serve the educational development needs of the region by providing education research in usable forms in order to promote school improvement and academic achievement and to correct educational deficiencies;
(F) facilitate communication between educational experts, school officials, and teachers, parents, and librarians, to enable such individuals to assist schools to develop a plan to meet the National Education Goals;
(G) provide training in--

(i) the field of education research and related areas;
(ii) the use of new educational methods; and
(iii) the use of information-finding methods, practices, techniques, and products developed in connection with such training for which the regional educational laboratory may support internships and fellowships and provide stipends;

(H) use applied educational research to assist in solving site-specific problems and to assist in development activities;
(I) conduct applied research projects designed to serve the particular needs of the region only in the event that such quality applied research does not exist as determined by the regional education laboratory or the Department of Education;
(J) collaborate and coordinate services with other technical assistance providers funded by the Department of Education;
(K) provide support and technical assistance in--

✓ (i) replicating and adapting exemplary and promising practices;
(ii) the development of high-quality, challenging curriculum frameworks;
(iii) the development of valid, reliable assessments which are linked to State, local, or Bureau-funded content and student performance standards and reflect recent advances in the field of educational assessment;
(iv) the improvement of professional development strategies to assure that all teachers are prepared to teach a challenging curriculum;
(v) expanding and improving the use of technology in education to improve teaching and learning;

(vi) the development of alternatives for restructuring school finance systems to promote greater equity in the distribution of resources; and
(vii) the development of alternative administrative structures which are more conducive to planning, implementing, and sustaining school reform and improved educational outcomes; and

(L) bring teams of experts together to develop and implement school improvement plans and strategies.

(4) NETWORKING.--In order to improve the efficiency and effectiveness of the regional laboratories, the governing boards of the regional laboratories shall establish and maintain a network to--

(A) share information about the activities each laboratory is carrying out;
(B) plan joint activities that would meet the needs of multiple regions;
(C) create a strategic plan for the development of activities undertaken by the laboratories to reduce redundancy and increase collaboration and resource-sharing in such activities; and
(D) otherwise devise means by which the work of the individual laboratories could serve national, as well as regional, needs.

(5) ADDITIONAL DUTIES.--Each regional education laboratory receiving assistance under this subsection shall carry out the following activities:

(A) Collaborate with the Institutes established under section 931 in order to--

(i) maximize the use of research conducted through the Institutes in the work of such laboratory;
(ii) keep the Institutes apprised of the work of the regional educational laboratory in the field; and
(iii) inform the Institutes about additional research needs identified in the field.

(B) Consult with the State educational agencies and library agencies in the region in developing the plan for serving the region.
(C) Develop strategies to utilize schools as critical components in reforming education and revitalizing rural communities in the United States.
(D) Report and disseminate information on overcoming the obstacles faced by rural educators and rural schools.
(E) Identify successful educational programs that have either been developed by such laboratory in carrying out such laboratory's functions or that have been developed or used by others within the region served by the laboratory and make such information available to the Secretary and the network of regional laboratories so that such programs may be considered for inclusion in the national education dissemination system.

-###-

PART C--NATIONAL RESEARCH INSTITUTES (cont.) PREV UP NEXT PART D--NATIONAL EDUCATION DISSEMINATION SYSTEM (cont.)

O

federal register

Monday
June 3, 1996

Part V

Department of Education

34 CFR Part 701

Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI)—Designation of Exemplary and Promising Programs; Proposed Rule

DEPARTMENT OF EDUCATION

34 CFR Part 701

RIN 1850–AA52

Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI)—Designation of Exemplary and Promising Programs

AGENCY: Department of Education.

ACTION: Notice of proposed rulemaking (NPRM).

SUMMARY: The Assistant Secretary is developing these standards pursuant to the Office of Educational Research and Improvement's authorizing legislation, the "Educational Research, Development, Dissemination, and Improvement Act of 1994." The major purpose of these standards is to provide quality assurance that programs designated by the Department of Education as either exemplary or promising have met criteria that will allow educators, professional organizations, and others to use these programs with confidence.

DATES: Comments must be received on or before August 2, 1996.

ADDRESSES: All comments concerning these proposed regulations should be addressed to Eve M. Bither, U.S. Department of Education, 555 New Jersey Avenue, N.W., Room 500, Washington, D.C. 20208–5530. Comments may also be sent through the Internet to: (Eve Bither@ed.gov).

Comments that concern information collection requirements should be sent to the Office of Management and Budget at the address listed in the Paperwork Reduction Act section of this preamble.

FOR FURTHER INFORMATION CONTACT: Sharon Bobbitt, U.S. Department of Education, 555 New Jersey Avenue, N.W., Room 508, Washington, D.C. 20208–5643. Telephone: (202) 219–2126. Internet: (Sharon Bobbitt@ed.gov).

Individuals who use a telecommunications device for the deaf (TDD) may call the Federal Information Relay Service (FIRS) at 1–800–877–8339 between 8 a.m. and 8 p.m., Eastern time, Monday through Friday.

SUPPLEMENTARY INFORMATION:

Background

On March 31, 1994, President Clinton signed Pub. L. 103–227, which includes Title IX, the "Educational Research, Development, Dissemination, and Improvement Act of 1994" (the "Act"). The Act restructured the Office of Educational Research and Improvement (OERI) and provided it with a broad mandate to conduct an array of research, development, dissemination, and improvement activities aimed at strengthening the education of all students.

Statutory Requirements

The legislation directed the Assistant Secretary to develop, in consultation with the National Educational Research Policy and Priorities Board, such standards as may be necessary to govern the conduct and evaluation of all research, development, and dissemination activities carried out by the Office to ensure that those activities meet the highest standards of professional excellence. The legislation requires that the standards be developed in three phases.

In the first phase, standards were promulgated to establish the peer review process and evaluation criteria to be used for the review of applications for grants and cooperative agreements and proposals for contracts. The final regulations setting out these standards were published in the Federal Register on September 14, 1995 (60 FR 47808). These proposed regulations address the second phase of development by establishing the criteria to be used in reviewing potentially exemplary and promising educational programs. The Assistant Secretary will publish at a later date additional proposed regulations for phase three of the standards, which are to govern evaluation of the performance of recipients of grants and contracts and cooperative agreements with OERI.

The OERI legislation requires that expert panels be established to review educational programs and recommend to the Secretary those programs that should be designated as exemplary or promising and disseminated through the Department's National Education Dissemination System. The legislation further requires the Assistant Secretary to develop standards that describe the procedures the panels will use in reviewing the educational programs. Section 941(a)(3) of the legislation broadly defines "educational program" to include "educational polices, research findings, practices and products." Educational programs may range in size and complexity from an individual instructional program—such as an elementary school science program—to a comprehensive reform initiative involving multiple goals and participants. Programs at all levels of education—preschool, elementary, secondary, and postsecondary—are eligible for consideration.

The Act also requires that the Assistant Secretary review the procedures utilized by the National Institutes of Health (NIH), the National Science Foundation (NSF), and other Federal departments or agencies engaged in research and development and actively solicit recommendations from research organizations and members of the general public.

In developing the review and evaluation procedures for the proposed standards, OERI has reviewed and considered dissemination practices and procedures used for identifying promising and exemplary programs by various foundations, research organizations, associations, and Federal agencies including NIH, NSF, the Federal Emergency Management Agency, the Department of Health and Human Services, and the National Endowment for the Arts. OERI adapted these review and evaluation procedures as appropriate.

Proposed Standards

The proposed standards have been developed by the Assistant Secretary in consultation with the Board. The standards proposed in this NPRM—

- Require that expert panels be established to review educational programs and recommend to the Secretary those programs that should be designated as exemplary or promising and disseminated through the Department's National Education Dissemination System; and
- Establish a process that panels will use to review and evaluate educational programs and determine which programs to recommend to the Secretary for designation as exemplary or promising.

Educational programs may be submitted at any time for consideration for designation as exemplary or promising. In addition, the Assistant Secretary will periodically establish and announce in the Federal Register specific topic areas of high priority for which programs will be invited or sought out. The legislation also provides that the Secretary may identify educational programs for the panels to review.

Educational program submissions may include, as evidence of the effectiveness of the program, a range of assessments, evaluative information from users, and other objective performance indicators that are appropriate to the program. The legislation ensures that a panel may not eliminate any program from consideration based on the lack of one type of supporting data such as test scores.

A standing group of experts, which will include teachers and others, will be appointed by the Assistant Secretary as appropriate. From that group, the Assistant Secretary will select members who have relevant knowledge and experience in specific topic areas to form expert panels to review programs in accordance with the criteria in these proposed regulations.

In determining whether an educational program should be recommended as exemplary or promising, the panel is required by the legislation to consider (a) whether, based on empirical data, the program is effective and should be designated as exemplary, or (b) whether there is sufficient evidence to demonstrate that the program shows promise for improving student achievement and should be designated as promising. These proposed regulations require a panel to evaluate whether a program has met all of the criteria of educational effectiveness set forth in Subpart C of these proposed regulations. A panel may determine that a program shows promise for improving student achievement and recommend that the program be designated as promising if the program has met all of the criteria with respect to one context, or with one population. A panel may determine that a program is effective and recommend that the program be designated as exemplary if the program has met all of the criteria with respect to multiple contexts, or with multiple populations.

Use of these criteria for evaluating programs will ensure that programs disseminated by the Department are high-quality, research-based programs that have provided evidence indicating they have improved teaching or learning or both. The Department's dissemination system is designed to make programs available to the public as quickly as possible. The system will enable the Department to respond to all forms of requests for information and assistance, and to support the applications of research and best practice. The system will use electronic networking and the capabilities of:

National Research Institutes;
Educational Resources Information Center (ERIC);
Regional Educational Laboratories;
Department-supported technical assistance providers;
National Library of Education; and
Other public and private nonprofit entities, including education associations and networks.

Prior to the adoption of these standards, exemplary programs were validated by the Department's Program Effectiveness Panel (PEP) and disseminated through the National Diffusion Network (NDN). With the adoption of these standards, the Department will recognize and disseminate promising educational programs in addition to exemplary programs.

Regulatory Flexibility Act Certification

The Secretary certifies that these proposed regulations would not have a significant economic impact on a substantial number of small entities.

The small entities that would be affected by these proposed regulations are small local educational agencies (LEAs) and private schools receiving Federal funds under this program. However, the regulations would not have a significant economic impact on the small LEAs and private schools affected, because the regulations would not impose excessive regulatory burdens or require unnecessary Federal supervision. The regulations would impose minimal requirements to ensure the proper expenditure of program funds.

Paperwork Reduction Act of 1995

Section 701.4 contains information collection requirements. As required by the Paperwork Reduction Act of 1995, the Department of Education will submit a copy of this section to the Office of Management and Budget (OMB) for its review. (44 U.S.C. 3504(h))

These regulations affect the following types of entities eligible to submit a program for review: Any public or private agency, organization or institution, or individual.

The public reporting burden is estimated to range from 2 to 6 hours for each program submitted for review. The actual burden will be determined by how much descriptive information about their program each entity wishes to provide.

Organizations and individuals desiring to submit comments on the information collection requirements should direct them to the Office of Information and Regulatory Affairs, OMB, Room 10235, New Executive Office Building, Washington, D.C. 20503; Attention: Wendy Taylor.

The Department considers comments by the public on this proposed collection of information in—

- Evaluating whether the proposed collection of information is necessary for the proper performance of the functions of the Department, including whether the information will have practical utility;
- Evaluating the accuracy of the Department's estimate of the burden of the proposed collection of information, including the validity of the methodology and assumptions used;
- Enhancing the quality, usefulness, and clarity of the information to be collected; and
- Minimizing the burden of the collection of information on those who are to respond, including through the use of appropriate automated, electronic, mechanical, or other technological collection techniques or other forms of information technology (e.g., permitting electronic submission of responses).

OMB is required to make a decision concerning the collection of information contained in these proposed regulations between 30 and 60 days after publication of this document in the Federal Register. Therefore, a comment to OMB is best assured of having its full effect if OMB receives it within 30 days of publication. This does not affect the deadline for the public to comment to the Department on the proposed regulations.

Invitation To Comment

Interested persons are invited to submit comments and recommendations regarding these proposed regulations.

All comments submitted in response to these proposed regulations will be available for public inspection, during and after the comment period, in Room 600, 555 New Jersey Avenue, NW., Washington, DC, between the hours of 8:30 a.m. and 4 p.m., Monday through Friday of each week except Federal holidays.

Assessment of Educational Impact

The Secretary particularly requests comments on whether the proposed regulations in this document would require transmission of information that is being gathered by or is available from any other agency or authority of the United States.

List of Subjects in 34 CFR Part 701

Education, Educational research, Reporting and recordkeeping requirements.

Dated: May 22, 1996.

(Catalog of Federal Domestic Assistance Number does not apply)

Sharon P. Robinson,

Assistant Secretary for Educational Research and Improvement.

The Secretary proposes to amend Chapter VII of Title 34 of the Code of Federal Regulations by adding a new Part 701 to read as follows:

PART 701—STANDARDS FOR CONDUCT AND EVALUATION OF ACTIVITIES CARRIED OUT BY THE OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT (OERI)—DESIGNATION OF EXEMPLARY AND PROMISING PROGRAMS

Subpart A—General

Sec.

701.1 What is the purpose of these standards?
701.2 What definitions apply?
701.3 What entity is eligible to submit a program for review?
701.4 What must an entity submit for review?

Subpart B—Selection of Panel Members

701.10 How are panels established?
701.11 Who may serve as a member of the standing group?
701.12 How is the membership of expert panels determined?

Subpart C—The Expert Panel Review Process

701.20 How does an expert panel evaluate programs?
701.21 What is the difference between an exemplary and a promising program?
701.22 What criteria are used to evaluate programs for exemplary or promising designation?

Authority: 20 U.S.C. 6011(i)

Subpart A—General

§ 701.1 What is the purpose of these standards?

(a) The standards in this part implement section 941(d) of the Educational Research, Development, Dissemination, and Improvement Act of 1994.

(b) These standards are intended to provide quality assurance that programs designated by the Department of Education as either exemplary or promising have met criteria that will allow educators, professional organizations, and others to use these programs with confidence.

(Authority: 20 U.S.C. 6011(i)(2)(E))

§ 701.2 What definitions apply?

Definition in the Educational Research, Development, Dissemination, and Improvement Act of 1994. The following term used in this part is defined in 20 U.S.C. 6041(a)(3):

Educational program

(Authority: 20 U.S.C. 6041(a)(3))

§ 701.3 What entity is eligible to submit a program for review?

Any public or private agency, organization, or institution, or an individual, may submit an educational program for review.

(Authority: 20 U.S.C. 6011(i)(2)(E))

§ 701.4 What must an entity submit for review?

(a) To have its educational program considered for designation as exemplary or promising, the eligible entity must submit to the Secretary a description of the program and a discussion of the program's educational effectiveness, responsive to the criteria in Subpart C, § 701.22.

(b) Information submitted must include, to the extent relevant to the particular program—

(1) A program abstract of 250 words or less;

(2) A description of the salient features of the program;

(3) A description of the program's philosophy and history;

(4) Site information, including demographics;

(5) A description of evaluation results;

(6) Funding and staffing information; and

(7) Organization name, address, telephone and fax numbers, e-mail address (if available), and contact person.

(Authority: 20 U.S.C. 6011(i)(2)(E))

Subpart B—Selection of Panel Members

§ 701.10 How are panels established?

(a) The Assistant Secretary, in consultation with the National Educational Research Policy and Priorities Board established under 20 U.S.C. 6021, establishes a standing group of educational experts. The Assistant Secretary may expand the membership of the standing group as necessary.

(b) The Assistant Secretary selects members from the standing group, based on their areas of expertise, to serve on expert panels in specific topic areas for the purpose of reviewing and evaluating educational programs and recommending, to the Secretary, those programs that should be designated as exemplary or promising.

(Authority: 20 U.S.C. 6011(i)(2)(E), 6041(d))

§ 701.11 Who may serve as a member of the standing group?

An individual may serve as a member of the standing group for the purpose of reviewing and evaluating educational programs for exemplary or promising status if that individual possesses two or more of the following qualifications:

(a) Demonstrated expertise and experience in one or more specific educational areas.

(b) Demonstrated expertise and experience across a broad range of educational policies and practices.

(c) Experience in evaluating educational programs.

(d) Experience or expertise in developing educational products.

(e) Current employment as a teacher, principal or other school-based or community-based professional (such as a guidance counselor, school media specialist, or health professional).

(Authority: 20 U.S.C. 6011(i)(2)(E), 6041(d))

§ 701.12 How is the membership of expert panels determined?

(a) For the review of each program, or group of programs, the Assistant Secretary establishes an expert panel comprised of individuals who are members of the standing group.

(b) In establishing the membership of each expert panel, the Assistant Secretary—

(1) Selects individuals who have in-depth knowledge of the subject area or content of the program or group of programs to be evaluated;

(2) Selects at least one current teacher, principal, or other school-based or community-based professional;

(3) Ensures that no more than one-third of the panel members are employees of the Federal Government; and

(4) Ensures that each panel member does not have a conflict of interest, as determined in accordance with paragraph (c) of this section, with respect to any educational program the panel member is asked to review.

(c) Panel members are considered employees of the Department for the purposes of conflicts of interest analysis and are subject to the provisions of 18 U.S.C. 208, 5 CFR 2635.502, and the Department's policies used to implement those provisions.

(Authority: 20 U.S.C. 6011(i)(2)(E), 6041(d))

Subpart C—The Expert Panel Review Process

§ 701.20 How does an expert panel evaluate programs?

(a) Each panel member shall—

(1) Independently review each program based on the criteria in § 701.22;

(2) Provide written comments based on an analysis of the strengths and weaknesses of the program according to the criteria;

(3) Participate in site visits if appropriate; and

(4) Participate in a meeting of the expert panel, if appropriate, to discuss the reviews.

(b) A panel may not eliminate an educational program from consideration based solely on the fact that the program does not have one specific type of supporting data, such as test scores.

(c) Each expert panel shall make a recommendation to the Secretary as to whether the program is exemplary, promising, or neither.

(Authority: 20 U.S.C.6011(i)(2)(E), 6041(d))

§ 701.21 What is the difference between an exemplary and a promising program?

(a) A panel may recommend to the Secretary that a program be designated as promising if the panel determines that the program has met each of the criteria of educational effectiveness in § 701.22 with respect to one context or one population.

(b) A panel may recommend to the Secretary that a program be designated as exemplary if the panel determines that the program has met each of the criteria of educational effectiveness in § 701.22 with respect to multiple contexts or multiple populations.

(Authority: 20 U.S.C.6011(i)(2)(E), 6041(d)(2))

§ 701.22 What criteria are used to evaluate programs for exemplary or promising designation?

In determining whether an educational program ("program" includes educational polices, research findings, practices and products) should be recommended as exemplary, promising, or neither, each expert panel shall consider the following criteria of educational effectiveness:

(a) *Evidence of success.* The expert panel considers—

(1) Whether, based on a range of assessments, information from users, or other indicators as appropriate, the program contributes to solving substantial or important problems in teaching or learning; and

(2) The extent to which—

(i) Program effects are beneficial to the populations for whom the program was designed; or

(ii) The product performs as expected for the educational consumers it was said to benefit.

(b) *Quality of the program.* The expert panel considers—

(1) Whether the program has clear goals, is based on sound research and practice, and incorporates accurate and up-to-date content;

(2) Whether the program represents a substantially improved alternative to existing options;

(3) The extent to which the program promotes equity and is free of bias based on race, gender, age, culture, ethnic origin, disability, or limited English proficiency status;

(4) Whether the program is based on high expectations for the success of all participants;

(5) Whether the program is appropriate to the target audiences; and

(6) The extent to which any materials associated with the program conform to accepted standards of technical quality.

(c) *Educational significance.* The expert panel considers—

(1) The extent to which the program has the potential to increase knowledge or understanding of educational problems, and issues, or effective strategies for teaching or learning; and

(2) Whether the program is described clearly enough so that it can be adapted or adopted in new sites.

(d) *Usefulness to others.* The expert panel considers—

(1) Whether the cost of the program (including money, staff time, and other required resources) is reasonable in light of expected benefits and compared to other alternatives; and

(2) Whether the program is available for use by others.

(Authority: 20 U.S.C. 6011(i)(2)(E), 6041(d)(2))

[FR Doc. 96–13801 Filed 5–31–96; 8:45 am]

BILLING CODE 4000–01–P

P

federal register

Monday
November 17, 1997

Part II

Department of Education

34 CFR Part 701

Office of Educational Research and Improvement (OERI) Conduct and Activities Evaluation Standards; Designation of Exemplary and Promising Programs; Final Rule

DEPARTMENT OF EDUCATION

34 CFR Part 701

RIN 1850–AA52

Standards for Conduct and Evaluation of Activities Carried out by the Office of Educational Research and Improvement (OERI)—Designation of Exemplary and Promising Programs

AGENCY: Department of Education.

ACTION: Final regulation.

SUMMARY: The Assistant Secretary for Educational Research and Improvement (OERI) is establishing final regulations pursuant to the "Educational Research, Development, Dissemination, and Improvement Act of 1994." The regulations are intended to provide quality assurance that programs designated by the Department of Education as either exemplary or promising have met criteria that will allow educators, professional organizations, and others to use these programs with confidence.

DATES: These regulations take effect December 17, 1997. However, affected parties do not have to comply with the information collection requirement in § 701.4 until the Department of Education publishes in the Federal Register notification of the compliance date and the control number assigned by the Office of Management and Budget (OMB) to this information collection requirement. Publication of the control number notifies the public that OMB has approved this information collection requirement under the Paperwork Reduction Act of 1995.

FOR FURTHER INFORMATION CONTACT: Stephen O'Brien, U.S. Department of Education, 555 New Jersey Avenue, NW, Room 502B, Washington, D.C. Telephone: (202) 219–2141. Internet: (Steve_O'Brien@ed.gov). Individuals who use a telecommunication device for the deaf (TDD) may call the Federal Information Relay Service (FIRS) at 1–800–877–8339 between 8 a.m. and 8 p.m., Eastern time, Monday through Friday.

Individuals with disabilities may obtain this document in an alternate format (e.g., Braille, large print, audiotape, or computer diskette) on request to the person listed in the preceding paragraph.

SUPPLEMENTARY INFORMATION: On March 31, 1994, President Clinton signed Public Law 103–227, which includes Title IX, the Educational Research, Development, Dissemination, and Improvement Act of 1994 (the Act). The Act restructured OERI and provided it with a broad mandate to conduct an array of research, development, dissemination, and improvement activities aimed at strengthening the education of all students.

The Act directed the Assistant Secretary to develop, in consultation with the National Educational Research Policy and Priorities Board (the Board), the highest standards of professional excellence necessary to govern the conduct and evaluation of all research, development, and dissemination activities carried out by the OERI. The legislation requires that the standards be developed in three phases.

In the first phase, standards were promulgated to establish the peer review process and evaluation criteria to be used for reviewing applications for grants and cooperative agreements and proposals for contracts. The Department published final regulations setting out these standards on September 14, 1995 (60 FR 47808). The regulations in this announcement address the second phase of development by establishing the criteria for panels to use in reviewing potentially exemplary and promising educational programs. The Assistant Secretary will later publish proposed regulations for phase three of the standards, which will govern how OERI evaluates performance of its recipients of grants, contracts, and cooperative agreements.

The OERI legislation requires that expert panels be established to review educational programs submitted by individuals or organizations. The legislation also provides that the Secretary may identify educational programs for the panels to review. The statute requires the panels to recommend to the Secretary those programs that should be designated as exemplary or promising and disseminated through the Department's National Education Dissemination System. The law requires that each panel consist of appropriately qualified experts and practitioners and requires the Secretary to develop standards that describe the procedures the panels will use in reviewing the educational programs. Section 941(a)(3) of the law broadly defines educational programs to include educational polices, research findings, practices, and products. Educational programs may range in size and complexity from an individual instructional program—such as an elementary school science program—to a comprehensive reform initiative involving multiple goals and participants. Programs at all levels of education—preschool, elementary, secondary, and postsecondary—are eligible for consideration.

In determining whether an educational program should be recommended as exemplary or promising, each panel is required by the Act to consider: (a) Whether, based on empirical data, the program is effective and should be designated as exemplary or (b) whether there is sufficient evidence to demonstrate that the program shows promise for improving student achievement and should be designated as promising. The Act expressly states that a panel shall not eliminate a program from consideration based on the lack of one type of supporting data such as test scores.

The evaluation process set forth in the final regulations will ensure that programs disseminated by the Department are high-quality, research-based programs that have provided evidence indicating they have improved teaching, learning, or both, or has demonstrated other worthy educational performance outcomes. The Department's dissemination system is designed to make information about these promising and exemplary programs available to the public as quickly as possible. The system will enable the Department to respond to all forms of requests for information and assistance, and to support the applications of research and best practice. The system will use electronic networking and the capabilities of:

—National Research Institutes;
—Educational Resources Information Center (ERIC);
—Regional Educational Laboratories;
—Department-supported dissemination and technical assistance providers;
—National Library of Education;
—Eisenhower Regional Consortia and Clearinghouse, and
—Other public and private nonprofit entities, including education associations and networks.

Until recently, the Department validated exemplary programs through its Program Effectiveness Panel (PEP) and disseminated them through the National Diffusion Network (NDN). Since this program no longer exists, with the adoption of these standards the Department will evaluate and disseminate promising educational programs in addition to exemplary programs. The Department will also work in partnership with constituency groups who have expertise in the specific topic areas represented by the expert panels to develop coordinated procedures to maximize their involvement in this work.

On June 3, 1996, the Secretary published a notice of proposed rulemaking (NPRM) for this part in the

Federal Register (61 FR 27990–27993). These final regulations contain three major changes from the NPRM. These changes are fully explained in the "Analysis of Comments and Changes" elsewhere in this preamble. The changes pertain to the standing panel; the distinction between "promising" and "exemplary"; and the factors listed under the criteria expert panels will use to evaluate programs.

Analysis of Comments and Changes

In response to the Secretary's invitation in the NPRM, seven parties submitted comments on the proposed regulations. This included comments from individual members of two pilot panels (math/science and gender equity) that were appointed by the Secretary to field test the expert panel process. In addition to the public comment, comments from the Board's Subcommittee on Standards are addressed as required by the legislation. The full Board approved the final regulations at a meeting on September 26, 1997. An analysis of the comments and of the changes in the regulations since publication of the NPRM follows.

Major issues are grouped according to subject with appropriate sections of the regulations referenced in parentheses. Technical and other minor changes—and suggested changes the Secretary is not legally authorized to make under the applicable statutory authority—are not addressed.

Eligibility (§ 701.3)

Comments: One commenter asked for clarification on who is eligible to submit educational programs for designation as promising or exemplary. Specifically, this commenter asked whether federally-funded entities, such as the Regional Laboratories, will be required to go through this process; whether local agencies that receive Federal funding through states, such as under Title I of the Elementary and Secondary Education Act (ESEA), may submit programs on their own; whether sponsors need to be invited to submit or may submit on their own initiative; and whether for-profit entities may submit.

Discussion: The law provides that "individuals" or "organizations" may submit educational programs for review. Since the law is silent on the specific nature of the organizations, the Secretary believes that for-profit agencies would be eligible to submit programs for review. With respect to the OERI-funded Regional Educational Laboratories, the law provides that the Secretary may identify those programs for panel review. In addition, the Secretary believes that the Laboratories could submit one or more of their programs on their own initiative. The question of whether local agencies that receive Federal funding through a State or Federal entity, such as under Title I of the ESEA, can submit on their own or must go through their funding agency, will be addressed in administrative guidance.

Changes: None.

Content of Submissions (§ 701.4)

Comments: Three commenters made suggestions about this section. Two commenters believed that requiring funding and staffing information was burdensome and not germane to the designation of a program as promising or exemplary. One commenter believed that this section should require program sponsors to submit specific materials related to content and methods. Another commenter believed that this section should include the requirement that the program include evidence of sustainability of improvement with targeted student populations.

Discussion: The Secretary believes that funding and staffing information should be included to help determine whether an educational program should be recommended as either exemplary or promising. The Secretary agrees that sponsors should be required to submit information or materials specific to content and methods, as available and appropriate. The Secretary believes that the evidence of sustainability of student improvement should be evaluated by peer reviewers in accordance with § 701.22.

Changes: Section 701.4(b)(7) has been renumbered as § 701.4(b)(8) and a new § 701.4(b)(7) has been added to include a provision for specific materials relevant to content and methods.

Procedures for Submitting Educational Programs (New § 701.5)

Comments: One commenter believed that the regulations should contain more specificity about the procedure for submitting programs to the expert panels. This commenter requested specifics on who receives the submissions and whether they may be submitted at any time or only on specific dates.

Discussion: The Secretary agrees that the general submission procedures should be included in regulations. A sponsor seeking the exemplary or promising designation for its educational program may submit its program at any time for consideration to the Assistant Secretary, who will assign the submitted program to the appropriate panel for review. The individual expert panels will set appropriate timelines for program submissions. In addition, the Assistant Secretary will periodically establish and announce in the **Federal Register** specific topic areas of high priority. Sponsors of educational programs in these areas will be invited to submit them for consideration.

Changes: A new § 701.5 has been added to include general procedures for submitting educational programs for review by an expert panel.

Establishment of Panels (§ 701.10)

Comments: The Board's Subcommittee on Standards recommended a change to the expert panel system. The Subcommittee thought that the structure of having members of the expert panels drawn from a separate standing panel of educational experts was an unwieldy, overly-complicated structure. The Board recommended that the expert panels be formed separately from a standing panel, which would instead provide an administrative oversight and monitoring function for the expert panels.

Discussion: The Secretary agrees that the expert panel should be formed separately from a standing panel. The Secretary will determine the feasibility of establishing a separate standing panel for the oversight and monitoring functions referred to by the Board—functions which are administrative in nature and could also be performed by OERI staff. Elimination of a reference to a standing panel in the regulations would not alter the composition and function of the expert panels as outlined in the NPRM.

Changes: Section 701.10(a) has been removed, § 701.10(b) has been revised, § 701.11 has been removed, § 701.12(a) has been revised, and § 701.12 has also been renumbered as § 701.11.

Panel Membership (§§ 701.11 and 701.12)

Comments: One commenter observed that §§ 701.11 and 701.12 in the NPRM did not explicitly state that those serving on the panels would represent both the community of practice and that of research. One commenter believed that each panel should include one or more members with evaluation expertise in order to help evaluate evidence of effectiveness.

Discussion: The Secretary agrees with these comments.

Changes: A sentence has been added at the end of the renumbered § 701.11(a) (formerly § 701.12(a)) stating that the membership of the expert panels will represent both the community of practice and the community of research. Additionally, §§ 701.11(b)(3) and

701.11(b)(4) have been renumbered as §§ 701.11(b)(4) and 701.11(b)(5), respectively; and a new provision for § 701.11(b)(3) has been added to include the selection of at least one individual with expertise in evaluating educational programs.

Difference Between Promising and Exemplary Programs (§ 701.21)

Comments: Five comments were received on the distinction between promising and exemplary programs. As proposed in § 701.21, the distinction was based upon the generalizability of the educational programs. Promising programs had to meet each of the criteria of educational effectiveness in § 701.22 (success, quality, educational significance, and usefulness to others) with respect to only one "context or population." Exemplary programs had to meet each of the criteria "with respect to multiple contexts or multiple populations."

Two commenters believed that the distinction should stay the way it was in the NPRM, although one of those suggested some clarifying language. However, three commenters questioned the distinction on the basis that it was too narrowly and artificially drawn and did not reflect the commonly understood meaning of the words "promising" and "exemplary." In this regard, one commenter believed that promising programs should not have to meet every criterion in § 701.22 at the same level as exemplary programs. Two commenters believed that promising programs should have to meet the criteria at the same level as exemplary, but that the evidence required of promising programs should be less stringent and that exemplary programs should be held to a higher standard of evidence.

Discussion: The Secretary agrees with those commenters who questioned the proposed distinction and advised OERI to give a more common sense meaning to the terms "promising" and "exemplary." The Secretary believes that the distinction between promising and exemplary programs specified in legislation is sufficient to cover these concerns. The Secretary relies upon the expert judgment of the expert panel reviewers in determining the nature and weight of evidence necessary to designate a program as either promising or exemplary, and in applying the criteria listed in § 701.22 in making this determination.

Changes: A revision has been made to the distinction between "promising" and "exemplary."

Criteria (§ 701.22)

Comments: Five commenters provided comments on this section and suggested revisions to either the wording of the criteria or to the content of the factors listed under each criterion or both. These comments included comments from one member of the math/science pilot panel and three members of the gender equity pilot panel. Although the math/science panel member did provide comments specific to the proposed criteria and factors, the consensus of this panel was that the expert panel process would be better served if each panel developed its own factors specific to the content or discipline or both under review by the individual panel. One commenter suggested that the word "replicability" would better capture the concept for the criterion entitled "usefulness to others." In addition, OERI's Board (The Subcommittee on Standards) thought that the regulations should be as simple as possible and should give the expert panels as much discretion as possible in evaluating programs submitted for review.

Discussion: In addition to the math/science and gender equity panels, the Secretary will establish pilot panels in technology and early reading in the next year. The Secretary has determined that until the work of all four pilot panels is concluded, the regulation should retain only the four criteria outlined in the NPRM in § 701.22 and allow each panel the flexibility to establish its own individual factors under each criterion that are specific to its content or discipline. The fact that the comments from the public suggested various changes to the factors underscores the desirability of this approach. While the final regulations will therefore no longer require the expert panels to apply the factors listed in the NPRM, the Secretary encourages each panel to look at these factors as suggested examples. The Secretary will review the factors developed by all of the panels to see if the criteria set forth in the final regulations need to be modified.

Changes: The factors specified under each of the four criteria have been eliminated and the criterion, "usefulness to others" has been changed to "replicability."

Assessment of Educational Impact

In the NPRM the Secretary requested comments on whether the proposed regulations would require transmission of information that is being gathered by or is available from any other agency or authority of the United States.

Based on the response to the NPRM and on its own review, the Department has determined that the regulations in this document do not require transmission of information that is being gathered by or is available from any other agency or authority of the United States.

Electronic Access to This Document

Anyone may view this document, as well as all other Department of Education documents published in the Federal Register, in text or portable document format (pdf) on the World Wide Web at either of the following sites:

http://ocfo.ed.gov/fedreg.htm
http://www.ed.gov/news.html

To use the pdf you must have the Adobe Acrobat Reader Program with Search, which is available free at either of the previous sites. If you have questions about using the pdf, call the U.S. Government Printing Office toll free at 1-888-293-6498.

Anyone may also view these documents in text copy only on an electronic bulletin board of the Department. Telephone: (202) 219-1511 or, toll free, 1-800-222-4922. The documents are located under Option G—Files/Announcements, Bulletins and Press Releases.

Note: The official version of this document is the document published in the Federal Register.

List of Subjects in 34 CFR Part 701

Education, Educational research, Reporting and recordkeeping requirements.

(Catalog of Federal Domestic Assistance Number does not apply)

Dated: November 11, 1997.

Ricky T. Takai,

Acting Assistant Secretary for Educational Research and Improvement.

The Secretary amends chapter VII of title 34 of the Code of Federal Regulations by adding a new part 701 to read as follows:

PART 701—STANDARDS FOR CONDUCT AND EVALUATION OF ACTIVITIES CARRIED OUT BY THE OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT (OERI)—DESIGNATION OF EXEMPLARY AND PROMISING PROGRAMS

Subpart A—General

Sec.
701.1 What is the purpose of these standards?
701.2 What definitions apply?
701.3 Who is eligible to submit an educational program for review?

701.4 What must a program sponsor submit for review?
701.5 What are the procedures for submitting an educational program for review by an expert panel?

Subpart B—Selection of Panel Members

701.10 How are panels established?
701.11 How is the membership of expert panels determined?

Subpart C—The Expert Panel Review Process

701.20 How does an expert panel evaluate programs?
701.21 What is the difference between an exemplary and a promising program?
701.22 What criteria are used to evaluate programs for exemplary or promising designation?

Authority: 20 U.S.C. 6011(i), unless otherwise noted.

Subpart A—General

§ 701.1 What is the purpose of these standards?

(a) The standards in this part implement section 941(d) of the Educational Research, Development, Dissemination, and Improvement Act of 1994.

(b) These standards are intended to provide quality assurance that educational programs designated by the U.S. Department of Education as either exemplary or promising have met criteria that will allow educators, professional organizations, and others to use these programs with confidence.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.2 What definitions apply?

The following definitions apply to this part:

Assistant Secretary means the Assistant Secretary for the Office of Educational Research and Improvement.

Educational programs mean educational policies, research findings, practices, and products.

Program sponsor means a party submitting an educational program for designation by the Secretary as either promising or exemplary.

Secretary means the Secretary of the Department of Education or an official or employee of the Department acting for the Secretary under a delegation of authority.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.3 Who is eligible to submit an educational program for review?

Any public or private agency, organization or institution, or an individual may submit an educational program for review.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.4 What must a program sponsor submit for review?

(a) To have an educational program considered for designation as exemplary or promising, a sponsor must submit to the Secretary a description of the program, program materials, and a discussion of the program that is responsive to the criteria in § 701.22.

(b) Information submitted must include, to the extent relevant to the particular program,—

(1) A program abstract of 250 words or less;

(2) A description of the salient features of the program;

(3) A description of the program's philosophy and history;

(4) Site information, including demographics;

(5) A description of evaluation results;

(6) Funding and staffing information;

(7) Specific materials relevant to content and methods, as appropriate; and

(8) Organization name, address, telephone and fax numbers, e-mail address (if available), and contact person.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.5 What are the procedures for submitting an educational program for review by an expert panel?

(a) An applicant seeking the exemplary or promising designation for its educational program may submit its program at any time for consideration to the Assistant Secretary, who will assign the submitted program to the appropriate expert panel for review.

(b) The Assistant Secretary will periodically establish and announce in the Federal Register specific topic areas of high priority. Sponsors of educational programs in these areas will be invited to submit their programs for consideration.

(c) The individual expert panels will set appropriate timelines for reviewing program submissions.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

Subpart B—Selection of Panel Members

§ 701.10 How are panels established?

The Assistant Secretary selects individuals, based on their areas of expertise, to serve on expert panels in specific topic areas for the purpose of reviewing and evaluating educational programs and recommending, to the Secretary, those programs that should be designated as exemplary or promising.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.11 How is the membership of expert panels determined?

(a) For the review of each program or group of programs, the Assistant Secretary establishes an expert panel. The membership of the expert panels will represent both the community of practice and the community of research.

(b) In establishing the membership of each expert panel, the Assistant Secretary—

(1) Selects individuals who have in-depth knowledge of the subject area or content of the program or group of programs to be evaluated;

(2) Selects at least one current teacher, principal, or other school-based or community-based professional;

(3) Selects at least one individual with expertise in evaluating educational programs;

(4) Ensures that no more than one-third of the panel members are employees of the Federal Government; and

(5) Ensures that each panel member does not have a conflict of interest, as determined in accordance with paragraph (c) of this section, with respect to any educational program the panel member is asked to review.

(c) Panel members are considered employees of the U.S. Department for the purposes of conflicts of interest analysis and are subject to the provisions of 18 U.S.C. 208, 5 CFR 2635.502, and the Department's policies used to implement those provisions.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

Subpart C—The Expert Panel Review Process

§ 701.20 How does an expert panel evaluate programs?

(a) Each panel member shall—

(1) Independently review each program based on the criteria in § 701.22;

(2) Provide written comments based on an analysis of the strengths and weaknesses of the program according to the criteria;

(3) Participate in site visits or other verification activities, if appropriate; and

(4) Participate in a meeting of the expert panel, if appropriate, to discuss the reviews.

(b) A panel may not eliminate an educational program from consideration based solely on the fact that the program does not have one specific type of supporting data, such as test scores.

(c) Each expert panel shall make a recommendation to the Secretary as to whether the program is exemplary, promising, or neither.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.21 What is the difference between an exemplary and a promising program?

(a) In determining whether an educational program should be recommended as exemplary or promising, the panel shall consider—

(1) Whether, based on empirical data, the program is effective and should be designated as exemplary; or

(2) Whether there is sufficient evidence to demonstrate that the program shows promise for improving student achievement and should be designated as promising.

(b) The Secretary relies upon the judgment and expertise of peer reviewers, as established in § 701.11, to determine the nature and extent of evidence required to distinguish between promising and exemplary programs and to apply the four criteria established in § 701.22, and their own individual factors under each criterion in making this determination.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

§ 701.22 What criteria are used to evaluate programs for exemplary or promising designation?

The Secretary establishes the following evaluation criteria for expert panels to use in determining whether an educational program should be recommended as exemplary, promising, or neither:

(a) Evidence of success.

(b) Quality of the program.

(c) Educational significance.

(d) Replicability.

(Authority: 20 U.S.C. 6011(i)(2)(B)(iii) and (E), 6041(d))

[FR Doc. 97-30051 Filed 11-14-97; 8:45 am]

BILLING CODE 4000-01-P

VOLUME II

Section V. Conceptual and Planning Papers

(in Chronological Order):

List of Papers and Publications Related to the System of Expert Panels	**Q**
Developing and Operating Expert Panels: A Key Leadership and Coordination Role for OERI -- Susan Klein, OERI (Jan. 16, 1995)	**R**
Leadership in Developing a Nation-wide "Findbest" Education R&D Evaluation System: A New OERI Responsibility -- Susan Klein, OERI (Jan. 29, 1996)	**S**
Technical Assistance and the Creation of Educational Knowledge -- Brenda J. Turnbull, Policy Studies Associates (March 1996)	**T**
OERI Leadership Efforts in Designing Approaches to Identify and Share Promising and Exemplary Products, Programs and Practices: Presession to the 1996 Annual Convention of the American Educational Research Association, New York City, April 8, 1996 Meeting notes by Allen Schmieder, OERI	**U**
Responses to: Options in Discussing Distinctions Between Promising and Exemplary Products, Programs, Practices, Policies and Research Findings for a System of Expert Panels -- Susan Klein (Nov. 19, 1996)	**V**
Expert Panels for Promising and Exemplary Innovations: A "Fine Idea From the Feds" by Peter Seidman, NCRVE Centerwork: National Center for Research in Vocational Education, Univ. of California, Berkeley (Winter 1997)	**W**
"Design Competitions: A Proposal for a New Federal Role in Educational Research and Development" by Robert E. Slavin, *Educational Researcher* (Jan./Feb. 1997)	**X**
"Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop, and Share Promising and Exemplary Products and Programs" by Susan S. Klein, *Educational Researcher* (Aug./Sept. 1997)	**Y**
The Evolving System of Expert Panels, A Presession on April 13, 1998 at the American Educational Research Association Conference, San Diego, Sponsored by the Office of Educational Research and Improvement, OERI, U.S. Department of Education. (Prepared by Chris Dwyer, Susan Klaiber and Peggy Simon, RMC)	**Z**
The Expert Panel System is a Reasonable Way to Identify Promising and Exemplary Programs --- A Response to Stanley Pogrow's Article, "What is an Exemplary Program and Why Should Anyone Care? A Reaction to Slavin and Klein." (Susan S. Klein, October 26, 1998).	**AA**

Q

Papers and Publications Related to the System of Expert Panels

The documents in this list are in addition to papers in this volume and the Dec. 1997 working papers volume on *Disseminating Promising and Exemplary Programs: Planning a System of Expert Panels.*

Papers on the Gender Equity Expert Panel and the Mathematics and Science Expert Panel

Formative Evaluation of the Pilot Expert Panel in Gender Equity "First Year Formative Evaluation: The Pilot Expert Panel in Gender Equity," Nov. 1996. Patricia B. Campbell, Campbell-Kibler Associates
Groton, MA

"Presenters Comments: Symposium on the Gender Equity Expert Panel", AERA Annual Meeting, Chicago, March 25, 1997.

"The Gender Equity Expert Panel Initiated by the U.S. Department of Education," by Susan Klein, Presented at the National Coalition for Women and Girls in Education Annual Conference, Maui, Hawaii, July 21, 1997.

"How the Gender Equity Expert Panel Can be a Gem" by Lois-ellin Datta, Presented at the National Coalition for Women and Girls in Education Annual Conference, Maui, Hawaii, July 21, 1997.

Formative Evaluation of the Pilot Expert Panel in Mathematics and Science Education Senta Raizen and Patti Bourexis, National Center for Improving Science Education, Washington, DC

"The Expert Panel System Procedural Guidebook - Based upon the Expert Panel in Mathematics and Science Education's Pilot Phase," Jan. 1998 Pelavin Research Institute, Washington, DC.

Earlier Related Papers and Publications

"Whose Knowledge is It?: Involving Teachers in the Generating and Using of Information on Educational Innovations", Patricia B. Campbell, Campbell-Kibler Associates, July, 1994

"A Matter of Consensus", Lois-ellin Datta, Datta Analysis, Aug. 1994

"Readiness for Change, Educational Innovations, and Educational Reform", Thomas E. Backer, Human Interaction Research Institute, Feb. 1995.

"Special Feature: Sharing the Best: Finding Better Ways for the Federal Government to Use Evaluation to Guide the Dissemination of Promising and Exemplary Education Solutions" *Evaluation and Program Planning*, July-Sept. 1993 (16-3) Pergamon Press, NY.

"Foreword" Michael Scriven

"Guest Editor's Introduction: Sharing the Best: Finding Better Ways for the Federal Government to Use Evaluation to Guide the Dissemination of Promising and Exemplary Education Solutions" Susan Shurberg Klein

"Best Wishes and Many Happy Returns: Some Federal Efforts in Recognizing the Best" Lois-ellin Datta

"Are There Better Ways to Identify and Share the Best Within and Among Federal Education Programs?" Susan Shurberg Klein

"Identifying the Best in Mathematics and Science in the U.S. Department of Education" Conrad G. Katzenmeyer

"Strategies to Share the Best in Postsecondary Education at the Fund for the Improvement of Postsecondary Education" Thomas G. Carroll

"Sharing the Best: The Dissemination System in Vocational and Adult Education" Joyce Fralic Cook

"Searching for the Best: The Chapter 1 Experience in the U.S. Department of Education" Elois M. Scott

"A Federal Perspective on Improving Practices, Programs, and Policies in Special Education" Martin Kaufman, Ellen Schiller, Beatrice Birman, Martha Coutinho

"Toward a Reciprocal Research, Development and Dissemination System: The Struggle for a New Paradigm in the Reauthorization of the Office of Research and Improvement" Laurence Peters

FEDERAL APPROACHES TO IDENTIFY AND SHARE THE BEST EDUCATION R&D SOLUTIONS
Highlights from the Special Feature in the July - Sept. 1993 *Evaluation and Program Planning*

Many who have participated in federal efforts to improve education believe that increased attention to identifying and sharing the best of our nation's education solutions is one of the most cost effective ways to substantially increase the positive impact of federal leadership and financial support. Since the 1970's there has been little systematic federal education leadership in identifying and sharing the best education solutions -- whether they were paid for by federal funds or by other governmental or non-governmental organizations. Figure 1 highlights current problems and future possibilities in ED efforts to maximize investments in what works at the project or solution level as they use four typical funding strategies.

The plans of U.S. Office of Education officials during the 1970s for a single unified system to identify and disseminate exemplary products, programs, and practices have never been fully implemented. Due to differences in federal education legislation and other causes, various offices in ED employ a variety of quality control strategies to guide the dissemination of promising and exemplary education solutions. Within ED, key senior staff involved in developing some of these evaluation approaches believe that the various offices have something to teach each other as well as others outside of education. As Dr. Michael Scriven has reminded us in his foreword, this special feature is one attempt to share this wisdom and encourage action. The articles describe how some federal offices are developing sensible procedures to use evaluation to identify and disseminate the best solutions for a variety of users. The authors believe that increased systematic planning is needed to maximize benefits related to improved use of clearly articulated (useful, feasible, accurate and proper) evaluation procedures to guide governmental decisions on what to disseminate. Such an orientation should also increase the cost effectiveness of governmental investments in R&D and dissemination. The articles by Drs. Datta and Klein provide insights on how strategic evaluation and dissemination orientations should lead to increased likelihood of achieving educational goals. However, they readily acknowledge that ED is far from ideal in identifying and sharing the best that the nation has to offer.

In the first article, "Best Wishes and Many Happy Returns: Some Federal Efforts in Recognizing the Best," Dr. Lois-ellin Datta, classifies ED' current efforts as being in the mainstream compared to related efforts by other federal agencies ranging from the U.S. General Accounting Office to the Department of Health and Human Services. This ranking is based on comparisons on: 1.) the extent of strategic planning and 2.) how the efforts dealt with the following five evaluation issues -- defining the universes; taking context into account; identifying what is necessary, sufficient and reproducible; saturation; and dilution. In the next context setting article, Dr. Susan Klein in the Office of Educational Research and Improvement (OERI) addresses the question "Are There Better Ways to Identify and Share the Best Within and Among Federal Education Programs?" In doing so, she provides a framework to understand ED efforts to identify and share the best of what works. Key aspects of this framework are included in Figure 1. Then Dr. Conrad Katzenmeyer, formerly with ED and now the head of the evaluation unit in the National Science Foundation, NSF, Office of Studies, Evaluation and Dissemination describes ED and cross-agency efforts to do this in the areas of mathematics and science education. Dr. Thomas Carroll shows how an entire ED office, The Fund for the Improvement of Postsecondary Education (FIPSE), integrates its R&D, evaluation and dissemination strategies to improve practices in postsecondary education. Dr. Joyce Cook, describes how a much larger Office of Adult and Vocational Education resorts to multiple often uncoordinated strategies to identify and share the best. Dr. Elois Scott describes how two ED offices use three different evaluation strategies to identify and share the best in the Chapter I compensatory education programs. Then Drs. Martin Kaufman and Ellen Schiller and colleagues with experience in the Office of Special Education Programs (OSEP) describe how they ensure that knowledge from R&D and practice is used by special educators. In doing so they explain why they have abandoned past strategies in favor of their new interactive directions. The issue ends with brief reactions to these federal approaches by Dr. Laurence Peters, formally Counsel to the Subcommittee on Select Education and Civil Rights in the U.S. House of Representatives and currently with OERI. Dr. Peters provides insights on how a new vision for a reciprocal research, development, and dissemination system has been guiding bills for the reauthorization of OERI.

For Additional Information Contact special feature editor, Dr. Susan S. Klein, Office of Educational Research and Improvement, U.S. Department of Education, 555 New Jersey Ave., N.W., Washington, D.C. 20208-5645, Tel. 202-219-2038, Fax 202-219-1407, E-mail sklein @inet.ed.gov

Figure 1
ED Funding Approaches: Current and Proposed

Current	Proposed Alternative Scenario
I. R&D ■ most emphasis on research and dissemination of research-related public information ■ little emphasis on development and replication or in identifying gaps in array of available solutions ■ little emphasis on identifying what is best from research or development	**I. New R&D Support Emphases** ■ more development of promising solutions to fill gaps ■ more comparative evaluations to learn about common practices that work across solutions and about gaps in types of solutions ■ more R&D to improve evaluation and dissemination methodologies
II. Demonstration Projects ■ most are quasi development, local service projects ■ little replication or demonstration of most of these projects outside of local area ■ all that receive project funding are assumed to merit the same dissemination support ■ there is no comparative evaluation of similar demonstration projects to determine merit for continued dissemination support ■ few qualify as exemplary using the current PEP	**II. Identification, Evaluation & Revision of Promising Solutions** ■ multiple strategies to identify the best from all sources ■ consumers obtain benefits from field testing promising solutions ■ developers obtain feedback for revisions ■ researchers and disseminators learn about what works from evaluators/users
III. Dissemination Efforts ■ most emphasis is on disseminating quality resources, bit by bit within individual projects ■ the variety of dissemination mechanisms managed by ED offices place little emphasis on supplying information on the best, either promising or exemplary or on supplying comparative information for consumers to make informed choices ■ even if certain ED offices have multiple evaluation and dissemination efforts, the coordination relating to identifying and sharing the best is weak ■ most dissemination mechanisms within ED offices and across ED are fragmented and uncoordinated.	**III. Expanded & Systematized Dissemination Support** ■ use comparative evaluations of promising solutions to identify those that are exemplary ■ provide most dissemination support for the best solutions--more for exemplary than for promising or high quality. ■ provide comparative consumer information on the solutions ■ provide dissemination and TA support to exemplary solutions as needed -- preferably in clusters with related exemplary solutions ■ support continued evaluation and revision of exemplary solutions to learn more about how they work in different contexts and with different users. ■ establish a national education dissemination system which improves the sharing of promising and exemplary R&D solutions across topics and constituencies
IV. Service or Categorical Funding ■ little integration with R&D ■ little focus on identifying or encouraging the use of the best R&D solutions	**IV. Enhanced Delivery of Services via R&D-based Solutions** ■ integrate with R&D ■ provide incentives to evaluate and use promising and exemplary solutions in the provision of services

R

Developing and Operating Expert Panels: A Key Leadership and Coordination Role for OERI

I. Overall Strategy

Background:
OERI has an excellent opportunity to provide leadership in key enduring topic areas to maximize the joint impact of the nation's research, development, and dissemination or reform assistance investments. OERI's new Institutes are creating R&D agendas and its Office of Reform Assistance and Dissemination has responsibility for operating multi-topic dissemination, assistance and improvement programs. Other offices in OERI such as the NCES and the new National Library of Education can also provide complementary support to organizing and implementing work according to topic areas.

This paper proposes to use a new provision in the OERI legislation to help OERI take advantage of this leadership opportunity in a meaningful way that will result in more constituent involvement and high quality research-based decisions to guide dissemination and R&D support within a year. This new provision is the creation of panels composed of appropriately qualified experts and practitioners (Section 941 (d) p.127). It is suggested that they be called **expert panels**. Expert refers to expertise in the topic area (not of a researcher versus a practitioner). Thus, each panel will be composed of federal staff, education practitioners, researchers, evaluators and disseminators who have indepth expertise in the topic of their panel--whether it be a broad topic such as school-wide improvement or a curriculum specific topic such as mathematics education.

The legislation allows for ⅓ of each panel to be composed of federal staff and we consider this a prime way to obtain meaningful cross ED and interagency involvement—as well as a powerful incentive for staff development. The federal government will be paying for these panels, convening them and designing them to fulfill federal needs to manage agency R&D, evaluation and dissemination efforts to improve education. This federal staff involvement is crucial because insights gained from participation in the panels and decisions made about its key responsibilities such as the selection of promising and exemplary programs should be incorporated into the operation of the improvement and service programs operated by the agencies as well as in the identification of gaps to guide the selection of new R&D. If foundations want to participate in work related to the panels, their experts could also be included as panel members.

To use these panels to facilitate broad-based and meaningful constituency involvement in federal decisions about what's best to disseminate or what research is most needed to advance improvement in a given topic area, each panel will have some core responsibilities in common. But since circumstances in each topic area vary, they will also have a great deal of freedom to perform other functions needed by the federal offices with responsibility for the topic area. They will have ample opportunity to influence the design of the core panel features as well as the unique activities of their own panel. Section II of this paper will describe initial thoughts on core and optional functions.

A New Vision for Expert Panels:
When people think about serving on federal panels they usually think about nationally prestigious advisory councils composed of 12-20 politically acceptable individuals with an elected or appointed chairperson, that meet a few times a year and issue broad consensus based reports or recommendations. Our concept of expert panels is quite different and intended to be flexible to meet the needs of the federal leaders in each topic area and to actively involve the constituency groups who care most about the topic. Although we envision that our expert panels would have opportunities to meet face-to-face--including at the meetings of various professional groups and in the Washington, DC area, most of their work would be communicated via phone, fax and e-mail and be year-round. We also assume that the panels could be easily expanded as they identify the need for additional specialized expertise to make high quality decisions and they might become fairly large with some of their work done in subgroups.

While there would be some reimbursement for panel expenses including travel to meetings, much of the panelists participation would be seen as a professional responsibility much like work for professional associations (or as peer reviewers for journals) . Whenever feasible, their national contributions would be seen as part of their regular job. Thus, they wouldn't receive double compensation from the government. Serving on expert panels would be a way for the individuals to continually keep up with the current developments in their field and to obtain prestige for being selected for this national leadership. They could also be called on for paid assistance from others as long as conflict of interest situations are avoided.

Terms would be multi-year to foster continuity. Individuals who participate actively would be invited to continue their participation. In line with newer philosophies of teamwork, we don't anticipate the selection of panel chairs, but the federal staff will serve as (or arrange for) panel facilitators and support staff and possibly even pay for support contractors do accomplish specific tasks needed by the panel[1]. The federal staff for each panel would also structure the activities so that the experts' valuable time will be focused on their tasks and on their decisions.

Establishment of Expert Panels:
Most of the Panels would be started in OERI, but could be initiated in other ED Offices. ORAD would provide basic conceptual guidance, frameworks, and core responsibilities and criteria for the panels and facilitate helping panels learn from each others' experiences and use common definitions so as not to confuse the public. The Office of the Assistant Secretary would reserve funds for the operation of the panels. These funds would be released when the Institutes or others submit acceptable plans for initial operation of their panel. ORAD would also work with the submitters in developing acceptable plans for the panels and be involved in their approval. In connection with its oversight of standards for the conduct and evaluation of research, the OERI National Educational Research Policies and Priorities Board (the Board) will be asked to review and approve the core criteria and procedures to be used by the expert panels.

II. Primary and Optional Functions of the Expert Panels in Each Topic Area

Distinction Between Primary and Optional Expert Panel Functions:
The OERI legislation specifies that the Expert Panels should play a major role in deciding what is promising and exemplary to guide future decisions to fund follow-up evaluation and dissemination/ implementation of promising and exemplary programs[2].

More specifically, the OERI legislation says the expert panels are to:

> *(A) evaluate educational programs that have been identified by the Secretary under subsection (c) or that have been submitted to the Secretary for such evaluation by some other individual or organization; and*
> *(B) recommend to the Secretary programs that should be designated as*

---

[1] Such tasks might include: arranging for travel and panel meetings; assistance in the identification of potentially promising products, programs or policies to accomplish specified goals; helping prepare documentation of submissions so that panelists will have verified and comparable information to support claims of advocates for the potentially promising product or program; assistance in obtaining the voluntary involvement of experts in the topic to augment the work of the actual panel members in refining and developing consensus-based selection criteria; design of comparative evaluation instruments and procedures to evaluate promising products & programs with similar purposes.

[2] "Programs" are defined broadly in the OERI legislation to *"include educational policies, research findings, practices and products"* p. 125.

> *exemplary or promising educational programs.*
> (Section 941(d) (1) p. 127).

It is essential that these panels be established as soon as possible because:

> *The Assistant Secretary may not designate a program as exemplary or promising unless a panel established under paragraph (1) has recommended that the program be so designated.*
> (Section 941 (d) (4) p. 127-8.)

However, many believe that organizing OERI work by topic area with an Expert Panel for each is so sensible that the functions of these expert panels could well be expanded to cover additional planning, decision-making, and constituent involvement needs in each area. This expansion of Expert Panel functions should be congruent with Congressional intent, because it should help OERI in its legislatively mandated R&D and dissemination coordination functions.[3] These optional functions may also help federal staff in each topic area avoid the unnecessary development of semi-duplicative patchwork advisory functions generally performed by ad hoc advisors on a one-time basis.

To the extent they are relevant and feasible, decisions by the expert panels should be considered in subsequent funding decisions by ED peer review panels as well as in funding guidelines for ED service oriented programs ranging from the support of professional development activities to federal support to states for early childhood services. (See Klein et. al. Evaluation and Program Planning, *Special Feature: Sharing the Best: Finding Better Ways for the Federal Government to Use Evaluation to Guide the Dissemination of Promising and Exemplary Education Solutions*, July-Sept. 1993 16 (3) p. 209-278.)

Initial Discussion of the Primary Functions:

- **Legislative Responsibilities:**

The Expert Panels primary functions in designating promising and exemplary programs are not an isolated mandated responsibility of the Office of Reform Assistance and Dissemination (ORAD). Indeed, the responsibility for identifying and sharing what works is an integral part of the legislative mandate for all of OERI and for each of the five Institutes. However, ORAD is the appropriate

---

[3] For example, OERI is to work with the rest of ED to improve the coordination of educational research, development and dissemination activities within the Department and federal government to minimize duplication and maximize value of the total federal investment. (p.97).

Office to assume leadership in organizing this important coordinating function. In addition to specifying the Expert Panels in the ORAD sections of the legislation, ORAD is to:

> *create a national system of dissemination, development, and educational improvement in order to create, adapt, identify, validate, and disseminate to educators, parents and policymakers those educational programs that have potential or have been shown to improve educational opportunities for all students;"* (p.125)

The OERI legislation further specifies multiple mechanisms to do this including:

- Developing processes to **identify** successful programs from other federal, multi-state, state, and local agencies, etc. for dissemination through the national educational system. (Section 941(c) p. 126.

- Developing and operating one or more **expert panels*** composed of federal and non federal R&D experts and practitioners) **to designate programs and products as promising and exemplary** (Section 941 (d) p.127). OERI is also required to describe procedures for this and related identification work and use peer review standards and procedures (p. 98).

- Supporting the **evaluation of programs** that have been identified as potentially worthy candidates for dissemination using various mechanisms (p. 127).

- Providing (along with others) **dissemination and implementation support** for promising and exemplary programs using multiple dissemination mechanisms in OERI and other places. (Section 941 (e) p. 128.

Some similar provisions are included in the Improving Americas Schools Act (IASA) which reauthorizes the Elementary and Secondary Education Act, particularly the section on the National Diffusion Network. Identification and replication of the best models, demonstration programs, set of principles or standards are also seen as important federal responsibilities in other ED legislation.

- **Advantages of Creating Expert Panels on Specific Topics to Identify and Share the Best:**

  - Most R&D and comparative evaluations of replicable programs need to be focused on specific topics to be meaningful. One of the weaknesses of the National Diffusion Network's, Program Effectiveness Panel (PEP) was that it was so general it was impossible for the panelists to examine the relative merit of a program compared to others designed to accomplish similar purposes. The fact that the panel members generally had no indepth expertise in the particular program topic they were randomly asked to review, also weakened their ability to make appropriate comparative judgements about educational significance. Relatedly, the panelists would not know if the content covered was accurate[4] and they would rarely ask questions about social fairness or equity[5]. Also, the incentives and identification procedures were so general that they didn't encourage active or a full range of excellent submissions in any topic area.

- Dissemination and implementation activities for promising and exemplary programs need to be provided according to topic areas as well as through general purpose dissemination efforts. This matrix strategy suggests that some of the activities are best done with a topic focus where others need to fit into general purpose systems and to be aware of how they will operate in larger contexts. (The ERIC system of topic specific clearinghouses operating within a larger system with coordinated activities and common procedures is a good example of this mix of functions.) Currently, the topic focus activities and the general purpose activities are not well coordinated among themselves and the dissemination R&D interface is particularly weak and unsystematic.

- Expert Panels on specific topics will provide many coordination and constituency building benefits. They will provide a meaningful way for:

  - ORAD offices, Institutes, NCES, other ED offices and education related programs in other agencies to work together to find the best in their respective domains and to share this with key constituents of all.

  - ORAD and all these other offices to involve key experts in key constituent groups in the:

---

[4] The panelists were not sufficiently knowledgeable in the content area to know if what was taught was accurate and current. They also had very little familiarity with the actual product or program--usually only a few pages of description in the PEP submission.

[5] In the late 1970's and early 1980's, the Joint Dissemination Review Panel, the predecessor of PEP started to ask for claims that the effects of the treatment didn't discriminate against groups on the basis of sex, race, or ethnic origin and to ask for assurances that the materials and programs were socially fair—non sexist, racist, fair to handicapped etc..

- identification of potentially promising and exemplary programs,
- development of criteria or standards (for authoritative national recognition and replication decisions)
- development of R&D agendas to advance progress in each of the topic areas covered by the expert panel.
- collection of information on user needs and desires including connections to relevant education reform efforts.
- developing and supporting strategies to evaluate programs so as to provide comparative information on their strengths and weaknesses for future dissemination/implementation support decisions and to provide information to guide consumer selection decisions.
- providing guidance on OERI supported dissemination as well as dissemination support from others such as associations, foundations, publishers, and states.

o Expert Panels while initially designed to help select promising and exemplary programs can also include recognition strategies and serve other related consultation purposes which will also improve the quality of the efforts to identify and share promising and exemplary programs. These options will be discussed more fully under optional expert panel responsibilities.

- **Examples of some initial procedures for the Expert Panels to use in carrying out their primary functions:**

o Add to core criteria developed by the new PEP/NDN as needed for each topic area expert panel.

The PEP criteria should be basically the same for promising and exemplary, except that in order to be judged exemplary evidence of replicability and effectiveness would need to be substantially greater than to be judged promising and previous low ratings on other criteria ranging from content accuracy to social fairness would have to be substantially improved.

Where recognition and evaluation efforts exist for individuals or organizations, in the topic area, consider using the same or related criteria or standards as for the more replication focused efforts associated with selecting promising and exemplary programs which can be used in a variety of contexts.

The redesign work for both the National Diffusion Network (See 10/28/94 draft by Farquhar) and the School Recognition Programs (See 11/18 draft by Demarest) contain some thoughtful discussions of criteria. It is hoped that ORAD can take the leadership and develop some core criteria for all expert panels to use as they make replication and recognition decisions. We anticipate that both replication and recognition decisions will still be made on the basis of the adequacy of the claims and evidence to support them. We also anticipate that in some cases, particularly for outcome criteria, there would probably be much in common among recognition and replication claims. Each expert panel would be free to augment the criteria as long as they stayed within legislative intent. Many panels may choose to use congruence with national or state standards[6] as part of their content related criteria.

Since decisions on follow-up federal support would differ greatly according to resources available for each expert panel and since these follow-up funding decisions may even be made by different agencies, most panels will probably decide to have a multi-step decision process. Thus, if an agency decides to fund a comparative evaluation, it could design it and then encourage the programs that were selected as promising to submit a statement that they would comply with the evaluation and revision requirements of this comparative evaluation. Similarly, this would give an agency or interagency group flexibility in supporting a dissemination strategy which "bundles" complementary promising or exemplary programs or which supports dissemination activities designed to help educators make wise selection decisions among different promising programs with similar types of outcomes.

o Focus on specific identification and search strategies that make sense for the particular topic area and build on what has already been done. These search strategies would focus on identifying potential promising or exemplary replicable programs. In some topic areas, associations or clearinghouses have developed databases or consumer reports on what's best in their area. Others have books on the topic or research syntheses or even standards which identify important criteria and concepts. These could be augmented as needed by publicizing criteria and by encouraging relevant groups to identify potential candidates as well as by encouraging self-nominations. (Additional help could come from general search and identification procedures through SEAs, NDN Facilitators, Regional Labs, etc.)

---

[6] In some cases the state standards may be even better than national standards, particularly if they are refinements of the national standards that had been developed earlier.

- Select those replicable programs that are promising on a periodic basis (possibly twice a year).
- Select those replicable programs that are exemplary at least once a year unless there are no candidates.
- Develop unique ways to ensure that the work of the expert panel will be used by various federal programs, in developing R&D agendas and in dissemination and service systems. (OERI would develop standard ways to connect with its multipurpose dissemination and reform programs).
- Help design comparative evaluations of promising programs with similar purposes.

Initial Discussion of Optional Functions: (Perhaps in coordination with other panels or groups)

- **Incorporate recognition activities in the topic area under the guidance of the redesigned recognition programs .** This would probably include examining and using research syntheses or supporting new ones or updates, doing consensus work on criteria where it may be important to tap national values or expert agreement on complicated principles or concepts.
- **Participate in ED or other agency design & review activities to decide on further evaluation, revision, or dissemination/TA support for promising programs.**
- **Participate in ED or other agency design & review activities to decide on further evaluation, dissemination/TA/ implementation support for**

exemplary programs. This may also include urging that programs that receive federal support for services use either promising or exemplary programs unless they provide adequate justification for other choices.

- **Coordinate needs assessments to learn what users want and need in a given topic area.**
- **Advise foundations and others on new areas within the topic that need additional external support**
- **Work with media and others through agencies and selected promising and exemplary programs or recognized organizations or individuals to publicize the work of the panel, OERI and cooperating agencies.**
- **As individuals, provide consultive help to others in the topic area as long as it is not a conflict of interest with panel responsibilities.** (The Expert Panel coordinator may even provide such lists to the public with panel members permission.)

III. Initial Challenges

Need to develop common terminology for describing what we are doing so that our terms will have the same meaning across OERI, the Department of Education, Other Federal Agencies and with our constituents throughout the nation.

Additionally, to the extent possible, these terms need to be defined so that they have a positive constructive meaning to most users. Many terms are defined in the OERI legislation and in (Klein, 1993, "Guest Editors's Introduction--Sharing the Best: Finding Better Ways for the Federal Government to Use Evaluation to Guide the Dissemination of Promising and Exemplary Education Solutions" *Evaluation and Program Planning.)*

An initial list of key terms may include:
- expert panel (on x topic) Some think expert has an elitist connotation, but others believe that it is appropriate just so it doesn't refer to one type of person such as a researcher, but not a practitioner. We need a name that will clearly indicate what the panel is to do, that is different from other government panels, that indicates national level prestige (but not federal control)

- promising
- exemplary
- replication[7]
- recognition
- adaptation
- program, product, practice, model
- dissemination, implementation, technical assistance
- evaluation

---

[7] **Replication Strategies** refer to:
Identifying, Designating, and Disseminating **Promising and Exemplary Programs** (products, policies, models that can be used in multiple contexts and result in similar desirable outcomes) They may be adapted to varying degrees but have substantial core features that make them somewhat different from other models in each replication.

Recognition Strategies refer to:
Identifying individuals or organizations (schools, school districts) which demonstrate a substantial number of desirable process and outcome principles or concepts that experts agree are qualities associated with success. (They are not based on replication because people and most organizations (perhaps excluding franchises) are not easily replicable. Demarest suggests using different terminology for stages of recognition instead of the promising and exemplary terminology used for replication. Recognition strategies also emphasize self-assessment and evaluation for diagnosis, planning and improvement.

Both Replication and Recognition Strategies need to identify concepts and principles of what's desirable to guide the selection of the best in the topic area. In most topic areas, both strategies are needed. This would certainly be true for expert panels focusing on what's best to facilitate school wide improvement. Thus, it makes sense to have a school or school district recognition program and school report cards. It also makes sense to examine the relative merits of school-wide improvement models such as "Success for All" and "Accelerated Schools". Similarly, the expert panel focusing on mathematics and science education should be particularly useful in identifying promising and exemplary instructional materials. And recognition activities may be particularly appropriate for identifying excellence in pre-service teacher training programs or institutions which ED or NSF may want to support through the Eisenhower State and NSF programs.

Need to develop strategies where potential participants will want to buy into the Expert Panel concepts and procedures and where they will have a substantial role in approving core criteria and procedures for all panels as well as for their own panel.
It is also likely that a management consultant can help design panel management and support system procedures to make the work productive and a positive experience for all the participants.

The selection of panel members is also critical. It needs to be designed so that hard working diverse group of experts[8] will be nominated (including via self-nominations) and selected. Also, the process of obtaining both federal and non-federal panelists needs to be done so that it enhances the prestige of serving on the panel without antagonizing people who might not be selected initially. The use of sub groups or informal liaison members might be used to identify new experts for full participation in the panel.

Relationship of Expert Panels to Redesigned and Continuing OERI R&D and Dissemination and Reform Improvement Efforts:
Many of OERI's major R&D and Dissemination Programs are Being Redesigned. There are many possibilities for them to play important roles in these Expert Panels. However, OERI staff will need to figure out how to maximize the benefits to all and make sure that the redesigns of these programs complement the expert panel concepts. It is also possible that the ongoing and redesign efforts may develop additional optional functions for the expert panels.

Thus, federal staff working on efforts such as these will need to be consulted:

- OERI Strategic Planning and Budgeting
- The Institutes with a particular focus on how the Expert Panels can help them develop their R&D agenda and plans for National R&D Centers
- NCES
- National Library of Education

---

[8] Diversity among the panelists will be needed to represent full coverage of areas of expertise needed by the panel related to: the range of subtopics in the area; R&D, evaluation and dissemination and practitioner skills, and adequate representation of women, minorities and other frequently under-represented groups who have expertise in social fairness issues.

- Goals 2000 State Support Efforts particularly how the work of the Expert Panels can help with the state reform plans, standards and curriculum frameworks

- Regional Educational Laboratories and other OERI and ED technical assistance providers

- Multi-topic Dissemination programs such as ERIC, NDN, and Recognition Programs. Coordination with the identification and validation functions of the Program Effectiveness Panel aspects of the NDN should be particularly helpful.

- Grants programs in OERI, ED and other education related agencies

Start-up Challenges:

- Allocating some funds to operate the expert panels and involve constituents as fully as possible.

- Going from a few topics and panels to many (This can be facilitated by designating the new NDN PEP panel as a multi-topic panel. Hopefully, the special topic panels would have more follow-up evaluation and dissemination funding than NDN. Thus submitters would prefer to be considered by the specific topic expert panels. At any rate, we would design the systems so that the government doesn't pay for reviews of the same program under more than one panel, unless there is some special reason to do so.)

- Going from pilot to predicable and consistent operations, while maintaining necessary flexibility to take advantage of opportunities, revise criteria and fund follow-up in different ways from panel to panel.

- Deciding on initial expert panel topics. Expert Panels would focus on topic areas where there are already some replicable programs and some consensus on key concepts an criteria in the topic area. Possibilities include:

  School wide improvement programs (Farquhar, Demarest, Anson, Paulu, Fox, CREATE-Sanders) NDN is already planning to focus on the priority of school wide projects for their 1995 developer demonstrator awards.

  Integrated Services or Early Childhood (Paulu, Karp, Demarest, Murphy)

  Math/Science /Conrad Katzenmeyer, NSF (Eve Bither, Carole Lacampagne, Pat Ross) (Maybe A. Dorsey, L. Levinson, other ORAD representatives) May use auspices of the National Science and Technology Council (NSTC). The Feb. 23-5, 1995 meeting may be a good place to announce this panel and perhaps solicit nominations of panelists.

Gender Equity (Klein, WEEA, WEEA Pub. Center/ OCR, OGC, Vocational Ed., FIPSE, At Risk Institute) May also help prepare for Beijing Conference.

Bilingual/ and or Multi-cultural (Mahoney, Gonzalez, Gil Garcia, Mack. Is OBEMLA Academic Excellence Program still in ESEA legislation?

Initial Budget Estimates for Expert Panels

Budget Philosophy to Support Expert Panels: OERI would pay only limited core support and provide the staff leadership to ensure some common purpose and similar principles and definitions across all panels.

- **Expert Panels will be used to manage R&D dissemination decision making in selected topic areas, thus they will not be separately budgeted competing programs that require substantial OERI funds.**

  Expert Panels should be viewed as an intelligent way of managing much of OERI and related federal agency work that is concerned with identifying and sharing the best R&D based knowledge and solutions in given areas. They would allow decisions to be based on systematic and authoritative advice from a variety of experts. Thus, they would be supported from a combination of salaries and expenses and programmatic funds that are allowed for the administration of legislatively mandated research and development, evaluation and service activities. In fact, some of the support activities for the expert panels such as identifying potentially promising solutions in a given topic area could be a part of the ongoing work of an R&D Center, Regional Educational Laboratory, National Diffusion Network, or technical assistance center responsibility. As these procurements are planned, consideration would need to be given to insure that they provide the types of services most needed by the expert panels to help make "authoritative" decisions to guide the nation. It is quite likely that the key staff involved in this work would also be either represented on the panels or asked to work with the panel in a collegial fashion.

- **Direct costs for operating these expert panels would be minimal. Much of the work would be done directly by federal staff or ongoing contractors and grantees as part of their normal responsibilities.**

  The federal government would be required to compensate the 2/3 of the panelists who are not already federal employees. Pending legal interpretation, I would assume that if some of these panelists are already receiving funding from the federal government for work related to their panel responsibilities, they could be compensated out of their regular grants or contracts. However, in some of these cases, the federal panel coordinators would need to pay for the non-federal panel members' travel and per diem to attend panel meetings. The government may also need to provide them with computer network links, if they don't already have necessary access to be a fully functioning member of the expert panel.

  It is estimated that in the initial year there would be a minimum of 3 two day meetings. Two would be in Washington, DC area, and one would be at a national meeting of a key associations related to the panel topic. In addition to their regular duties at this meeting the panel would interact with the association members and public interested in their topic.

◆ **Cost-sharing by the various federal participants would be the norm, rather than the exception. This should help R&D and dissemination coordination within the Department and other agencies as required by OERI legislation.**

These panels would have a great deal of flexibility to emphasize various aspects of their topic area and to be generally responsive to the needs and resources of the cooperating agencies. For example, in the mathematics and science area, if the most immediate need in NSF and ED is to make more systematic decisions about what promising and exemplary materials and related resources should be disseminated through the national education dissemination programs operated by OERI and other parts of ED (such as the Comprehensive Technical Assistance Centers), they should each contribute where they have access to the experts and information. Thus, NSF might well provide subject area experts (perhaps they would select experts who have had substantial experience in their peer reviews). OERI and ED might supply experts who have experience in disseminating mathematics and science resources, perhaps from the Eisenhower Clearinghouse and from special population programs such as the Office of Bilingual and Minority Language Affairs and the Women's Equity Program. The principle would be to the extent that a federal program office selects an expert (agreed to by the others to provide an appropriate balance to the panel), they would also pay the expert's expenses—if not already covered by a federal grant or contract.. Such flexibility would also help increase the numbers of experts as needed for the general panel and even subpanels on more specific topics, such as teaching algebra.

Agencies could also cover related assistance prior to panel decision-making and after the decisions have been made. Thus, NSF may pay for a separate effort to identify potentially promising resources that have not received NSF or ED funding—perhaps even from other countries. ED might support research and practice syntheses or survey its clearinghouses and technical assistance providers, state curriculum frameworks directors, recognition programs, etc. Such extra activities might help the panel learn about the current state of the art research and about what practitioners appear to want and need but have trouble finding in the way of resources to guide the search strategies and the fine tuning of the criteria. After resources have been selected as promising, the agencies would be expected to provide follow-up support for needed revisions and subsequent evaluations to compare the resource with others having similar purposes and to help pay for evidence that may help subsequent panel determinations of exemplary status.

Potential OERI Core Expenses in Year 1 of one Panel might look like this:

This estimate is based on the assumption that OERI will pay directly for 6 of 12 experts to attend 6 days of meetings and to spend an additional 10 days on work related to their panel responsibilities. It also acknowledges some potential expenses for OERI staff travel to the meeting away from the DC area and the potential need to pay for some general meeting related expenses.

The following core OERI budget estimate suggests that OERI might be able to operate an active and highly engaged panel for **under $50, 000 in direct expenses for year 1**. (OERI staff time and contributions from other agencies or of panelists working on other federal projects are not in this budget estimate, because this panel work would directly contribute to improving their major work responsibilities.)

- Travel and expenses for 2 days = $1000 times 3 trips = $3000 times 6 non-federal panel members = $12,000
- Daily Compensation $300 per day times 16 days = $4,800 times 6 people =$28,800
- Expenses for Meeting facilities, support services*, invited guests, etc. $1000 per day times 6 days = $6000
- Travel and Per Diem for OERI staff on the panel or working with it to attend the out-of town meeting. (Assume 3 OERI staff would travel at $1000 each person or $3000.)

Estimated Total = $49,800

---

* Bruce Barkley said that he would help us design the initial meetings so they run productively. He is already being paid by OERI so this will not be an additional core expense.

S

Draft January 29, 1996

Leadership in Developing A Nation-wide "Findbest" Education R&D Evaluation System: A New OERI Responsibility

by

Susan S. Klein, Office of Reform Assistance and Dissemination, Office of Educational Research and Improvement, U.S. Department of Education[1]

Do you agree with any or all of the following options?

A. "The education sector has no accepted norms or standards for evaluating innovations before they can be marketed—. We have more standards for marketing hamburger than we do for introducing changes in the schools." (Carnine, 1995, p.59) "Positive incentives to supply and demand knowledge about better practice are missing." (Canine, 1995, p.60).

B. Educational practitioners in the U.S. want to obtain authoritative comparative advice on the myriad products, programs and practices (developed and used by others) so that they can select those that are most likely to help them improve their services. However, experience indicates that they don't want to pay for this information directly and that it doesn't make sense for each local district and state education agency to develop its own consumer information system.

C. While the federal government and others have supported bits and pieces of what works identification and dissemination efforts, these activities have not been coordinated conceptually or operationally. Increased coverage and coordination will enhance the appropriate use of R&D in guiding decision making of practitioners and federal staff.

D. All of the above.

---

[1] Dr. Klein wrote this article in her private capacity. The ideas expressed are the author's and they do not necessarily reflect the position or policy of the U.S. Department of Education; official endorsement is neither implied, nor should it be inferred.

The proposed Findbest System is a response to the above views[2] of what's lacking and what's needed to help improve one key quality control aspect of the decentralized education system in the U.S. U.S. educators are in a position to benefit from the creativity of researchers and developers who have produced many resources to improve education. But educators face a formidable challenge in learning about and selecting the R&D-based resources that are most likely to help them, their students and colleagues. As we approach the year 2000, the federal government is in a unique position to help educators capitalize on this wealth to address current and emerging educational challenges.

The Office of Eduational Research and Improvement (OERI) in the U.S. Department of Education (ED) can lead in the development of this new nation-wide R&D evaluation system to serve two primary types of consumers. The Findbest System will: 1. help federal and other funders and producers of education resources identify areas where new products and programs are needed and where they can maximize investments in the best by providing support for further evaluation, development and dissemination of existing promising and exemplary resources. 2. help educators and learners make wise selections of the best R&D-based resources based on a greater understanding of their relative merits.

OERI is particularly well suited to provide this leadership now because it has a legislative mandate to do so (U.S. Congress, 1994 a) and because it can do it:

- fairly inexpensively using increasingly available technologies such as Internet and new approaches to systemic reform.
- by building on related public consumer-oriented R&D evaluation efforts of other federal agencies, foundations, and educators. (OERI is commissioning papers to examine other consumer-oriented review systems to learn what works and what's best to inform the public of their options in: mathematics and science education, other

---

[2] Some educators and policy makers have different views on how to improve education. They believe that educators chiefly need to identify useful principles to construct their own products, programs and practices. Others believe that the replication of externally produced innovations is more likely to be a process of adaptation than adoption; making the evaluations of others' experiences with innovations of negligible value to potential users. Others believe that in the U.S. decentralized education system there is little role for federal leadership and coordination of education consumer information and that if this is done it should be done entirely outside of government auspices, much like the *Consumer Report* model for purchasing televisions. We plan to take these views into consideration as we design the Findbest System.

education topics, federal agencies, and foundations and non-profits. This work will build on studies of the agricultural extension system (Klein, 1992 b), ED and other agencies (Klein, 1992 a; Klein, Ed., 1993; Barkdoll, 1996), foundations (Backer, 1995) and consensus development strategies (Datta, 1994).

- by advancing evaluation methods particularly related to quality control and consumer information prior to dissemination and implementation support.

Congress assigned this new leadership role to the Office of Educational Research and Improvement (OERI) in the U.S. Department of Education (ED) when OERI was reauthorized in 1994 under Title IX of the Goals 2000: Educate America Act. Under this 1994 OERI Reauthorization, OERI is to:

- develop standards to recommend the designation of promising and exemplary education programs (The initial version of the standards should be published for public comment by March 31, 1996. The legislation defines "programs" broadly to include products, practices, programs, policies and research results. These standards will complement the standards OERI has developed to review grant proposals and the standards it is developing to assess the performance of its own programs.)

- use one or more expert panels composed of experts and practitioners in specific topic-areas to make the above recommendations to the Secretary of Education. (Pilot expert panels in two topic areas of national importance are being developed in 1996.)

- have the OERI Institutes and Regional Educational Laboratories help in the identification of what is best in their areas of responsibility and use this increased knowledge of options in their specialty fields ~~to~~ help identify gaps that can be filled by federal and other support of R&D.

- use OERI's Office of Reform Assistance and Dissemination (ORAD) leadership responsibilities in coordinating and managing a National Education Dissemination System (including the National Library of Education) to share and obtain evaluation information related to the determination of promising and exemplary programs and to encourage the continued evaluation, improvement, dissemination and implementation of the programs deemed promising and exemplary.

- conduct research on dissemination by carrying out "a program of research on models for successful knowledge dissemination, and utilization, and strategies for reaching education policymakers, practitioners and others interested in education" (p. 126, Section 941 (b) (2) (F))

- help educators and members of their communities increase their abilities to make wise selections of R&D and contribute evaluations of what they use. (Teacher Demonstration Development Program provision.)

- coordinate these activities with other ED and related federal offices and with constituency groups such as foundations and associations who have specific interests in identifying and using the best in their areas of special interest. (This role of learning and sharing what works is part of the ED Strategic Plan (1994) and specifically mentioned in other ED legislation such as the 1994 Improving America's Schools Act (U.S. Congress, 1994 b). It is also a continued Congressional interest since educators want federal help in obtaining systematic and understandable consumer information on high quality alternatives. For example, the House Economic and Educational Opportunities Committee in the 104 th Congress asked the U.S. General Accounting Office (GAO) to obtain information on what works and what doesn't work in education. The resulting GAO report is, *Schools and Workplaces: An Overview of Successful and Unsuccessful Practices* (GAO/PEMD-95-28, 1995).

The 1994 OERI reauthorization legislation is specific about some aspects of this work to designate and share promising and exemplary products, programs and practices. Such specifics include: the panel appointing authority of the Assistant Secretary of OERI, the advisory role of the OERI Board, the need to learn from review experiences of other federal agencies, the federal versus non-federal staff composition of the expert panels, and the need to compensate non-federal panel members for their work. The legislation allows a great deal of flexibility in other aspects of this work.

The next two sections of this paper will discuss how the Findbest System can be designed to address it's two key purposes: 1. improving the federal leadership role in supporting and sharing what works and 2. helping consumers make wise decisions in selecting educational resources.

1. How is the proposed Findbest System an example of a new type of federal leadership strategy that can produce a more effective, consumer-focused Department of Education ?

In the proposed Findbest System, the federal government will play a facilitative leadership, partnership, community building role, rather than a dictatorial, regulatory control role. This would involve:

- **Using new leadership strategies that are congruent with recent government reinvention and business management principles.**
  These principles depend on high involvement and participation of all stakeholders, a fairly flat non-hierarchical organizational structure, meaningful rewards for all who need to be involved, the development of good knowledge-based communications and skills, and meaningful comparative evaluations of performance. (Lawler, 1991; Wohlstetter, 1993).

- **Building a dynamic nation-wide, topic-focused SYSTEM.**
  Developing a well understood and articulated but flexible system is a crucial part of the plan and it is congruent with comprehensive systemic reform objectives in education and other areas. A dynamic nation-wide topic-focused systems approach is needed because:

  - It is more likely that OERI will be able to carry out its leadership responsibilities in identifying and sharing what works if it helps create a dynamic collaborative nation-wide system using common principles and procedures, than if it focuses its attention on isolated functions or parts. Thus, it will be will be necessary to use many of the same core criteria and procedures among the expert panels in the different topic areas so that users can expect consistent high quality information from each panel. As the professional development and incentive elements of this system are developed, they too will need to be coordinated with the panels and each other. For example, pre-service educators might teach their students to evaluate potential materials and to submit these evaluations to the overall system manager responsible for the appropriate topic. Where feasible, multiple federal agencies may want to provide incentives by supporting comparative evaluations of promising programs or by paying for the dissemination and marketing of useful bundles of exemplary programs.

  - Pieces of consumer-oriented evaluation activities are place, but they are not connected with each other and they are often one-shot efforts that are not updated as new resources and evidence becomes available. They do not cover all nationally important topic areas and potential users generally do not know about their availability or quality. There has been relatively little attention to learning from them so that future consumer-oriented evaluations will be improved. It is also very difficult for consumers to check information from more than one review effort. For example if users want to learn what works well for bilingual students in mathematics they will need to check various databases with some connection to mathematics and others with some connection to bilingual education. Even if they are able to find some information in both sources, the information will not be compatible and the review criteria and designations of quality (such as successful, promising, validated, or exemplary) will not be consistent. The federal government is in a position to provide leadership; linkage, and sustained support to build on and learn from these short term efforts and to develop a system that will involve all stakeholders in meaningful ways.

Examples of parts of this consumer-oriented evaluation system already in place include: Educational Products Information Exchange (EPIE), a non-profit consumer information organization; the Department of Education's Program Effectiveness Panel and related National Diffusion Network activities; work of the Regional Educational Laboratories in identifying promising mathematics and science education programs; ERIC and other Clearinghouse efforts to respond to requests to share what they know about the best in specific areas; and reports and books by experts in particular areas such as Heather Weiss (1995) in early childhood education. Many of these books describe the research-based principles and other criteria they used to select their best bets and sometimes they compare the relative strengths and weaknesses of those they have selected.

- A coordinated evaluation system would facilitate the development of a user feedback system that would help educators share information on how promising and exemplary programs have worked for them. It is expected that as teachers contribute their own evaluations to the system, they will learn more about what is available and become committed users and supporters of the system. (See an OERI commissioned paper by Patricia B. Campbell, 1994 for an initial discussion of such a system in "Whose Knowledge is It?: Involving Teachers in the Generating and Using of Information on Educational Innovations")

- As a coherent well understood R&D evaluation system, Findbest would be an integral part of the National Education Dissemination System (NEDS) as well as a system that could be coordinated with related systemic education reform efforts. It would also help OERI's National Research Institutes as they develop and implement their R&D agendas.

- A nation-wide system makes sense in the U.S. because:
  1) There are many potentially promising and exemplary resources that could benefit a variety of users across the nation; 2) It is increasingly feasible for a system to operate from one location and communicate fairly easily with stakeholders across the nation using toll-free telephone services, faxes, audio and video tapes, and of course, Internet. 3) It makes little sense for each school, school

district or state to create their own system to identify and review materials in a wide variety of content areas.

Nation-wide is used instead of federal or national. Nation-wide means that the system is intended to serve the entire nation, but not be totally controlled by a national organization. Instead, this new system will be dependent upon the participation of many national and other organizations and on individuals with nationally recognized expertise in the specific panel topic areas. The system may also acquire information on international resources and be used by educators in other countries, but it will be primarily intended for educators in the U.S. and only U.S. educators will participate as official reviewers/panel members. Unlike many other federal education improvement efforts, it will not be necessary for activities of this system to be filtered through local or state education agencies, although if these agencies want to share suggestions on promising and exemplary programs or participate in the evaluation and dissemination activities they will be most welcome.

- The topic focus of each expert panel is essential to facilitate meaningful comparisons of resources intended for similar purposes such as eliminating drug and violence problems in schools. It will also facilitate involvement of interested stakeholders such as professional associations which focus on the topic area and educators who need topic focused resources. While it is expected that there would be substantial similarities among panels such as the use of core criteria, each panel would be encouraged to use additional topic specific criteria and to take advantage of opportunities in their topic areas and to make meaningful comparisons and connections among the promising and exemplary solutions in their domain. While there is a substantial federal focus on providing technical assistance and reform services at the regional, multi-state, state and local district levels, only a few federal efforts have had an inter and intra agency focus on specific topic areas of national interest. Some of these exceptions have included the collaborative work in mathematics and science education and of some topic focused research centers and information clearinghouses. As with the ERIC clearinghouses, expert panel topics may be selected based on curriculum content (the arts), population groups (Limited English Proficient students), or processes (school-wide improvement efforts).

Topic focused expert panels will be composed of individuals with expertise in the topic such as teachers, researchers, evaluators, disseminators, producers and administrators. These practitioner, researcher and dissemination experts in each specific topic area (rather than generalists) would decide on promising and exemplary status of the solutions from all sources--not just those that received previous federal funding or that were seeking current federal support. These experts would be well-linked with their respective topic focused "learning communities" and they will contribute to R&D agendas as well as consumer information and technical assistance services in the National Education Dissemination System.

The initial pilot expert panels are in the areas of 1. gender equity and 2. mathematics and science education. Ideas that these panels are considering include using the initial panel members in an advisory capacity and as chairs of sub-panels in their areas of specific expertise such as "history of science" or "middle-school mathematics instruction". In this way they could act somewhat as journal editors and help recruit sub-panels members to involve more people, to identify likely promising and exemplary programs, and to conduct the necessary numbers of high quality reviews.

- **Collaborating within the federal sector and with those in other public and private roles to 1) operate the Findbest System and 2) provide incentives for participating.**

  The Findbest System would provide a clear, consistent and meaningful purpose for collaboration among federal offices responsible for similar types of educational improvements. If the system becomes sufficiently well respected, foundations other governmental agencies, foundations, professional associations and publishers will want to participate in advising, using and contributing resources to support the decisions of the expert panels.

  1) Operational collaboration would be based on fostering the most appropriate types of participation. Thus, a topic focused resource and technical assistance center might follow the model of the Women's Educational Equity Resource Center which is managing the pilot expert panel in gender equity and also plans to prepare consumer reports based on panel designations. This panel will also involve and try to serve the needs of ED's Office for Civil Rights which wants to be able to share successful practices with entities which are having compliance problems in areas such as sexual harassment, with the Office of Special Education related to the needs of disabled women and girls, and with the Women's Bureau and the ED Offices concerned with promoting gender equity in school to work and vocational and technical education.

  2) Federal and other sponsor incentives would include:

  - Funding promising programs, products or practices for revisions and further comparative evaluations. For example, the federal government or others could fund the comparative evaluations of solutions for similar challenges that an expert panel deemed promising. This would also be a way to support the continued and expanded use of the promising program. The supporting agencies may also establish procedures to obtain continuous systematic user feedback on promising and exemplary solutions.

- Providing consumer information on the promising and exemplary programs via existing personal, print, audio-video and computer-based information providers as well as from R&D databases. (This "routine" incentive would be intended to help potential users make informed selection decisions, but it would also provide some recognition for the developers and producers of the programs because they would be cited in these descriptions.)

- Providing special help and attention to increase the public's awareness of the value of the **exemplary** products, programs and practices. For example, the federal government might fund demonstrations, and dissemination of consumer information on clusters of exemplary R&D based solutions[3] or when agencies fund other professional development or student services grants, they may include incentives for grantees who select from among the exemplary resources.

- Funding research and development to fill needed gaps in the array of solutions in the topic areas. (The OERI Research priorities which guide the work of the Institutes and their R&D Centers and the specialty area work of the Regional Educational Laboratories could be guided in part by the work of the expert panels and staff from these entities would be officially connected with their respective expert panels.)

- **Strong and prolonged federal support for the operation of a long term dynamic system composed of individual panels in as many topic areas of interest to the American public as possible.**
  If this system works as intended, it will influence other federal education activities such as encouraging more federal funding of evaluations of replicable products, programs and practices and making it possible to fund demonstration programs that have been designated as at least promising and maybe exemplary. To do so the Findbest System will need a long-term core operational budget and the ability to have other federal offices and foundations pay for work that is within their purview such as

---

[3] Comparative information would be provided on items in each cluster so that users could decide if they want to select one or several to best fit their needs.

for subpanel members, comparative evaluations of promising programs, consumer reports, etc.

2. **How will the Findbest System help educators and learners make wise selections of the best R&D-based resources?**

Educators and students themselves are continually asking what is best to do or learn "whatever". Often technical assistance providers don't have knowledge of the options or their relative strengths and weaknesses. The findbest question is fundamentally difficult because the answers may change as R&D-based products are developed, improved or tested and used if different ways. In designing the findbest system to help educators and learners, OERI will need to provide leadership in:

- **Creating new consumer-oriented evaluation criteria and procedures.** Consumer-oriented evaluation is an emerging concept—perhaps one that will be renamed. It is different from other terms such as client-focused or stakeholder or empowerment evaluation, but includes some of the principles of each of these approaches. Consumer-oriented evaluations have the following features:

  - *Consumer* is defined as the entire public that might be interested in learning about the merit of a product, program or practice . Governmental agencies, foundations and non-profits that operate to advance the public good and (in some cases to protect the public from harm) are important users of consumer-oriented evaluation strategies. If they are focusing on the consumer service (ED, National Institutes of Health, Agency for Health Care Policy and Research) more than the consumer-protection aspect (Federal Drug Administration), they will need to evaluate all plausible options to accomplish their desired purposes before they decide to disseminate the best of what is available to their publics.

  - Since there are many products, programs and practices with similar purposes, consumers will want to know about their comparative merit on a range of criteria that are likely to be important to their own decision-making. In the *New Directions for Evaluation*, winter 1995 issue on "Reasoning in Evaluation: Inferential Links and Leaps" edited by Deborah Fournier, she and others discuss various aspects of the working logic of a consumer approach such as asking, "Is X a good/less good one of its type"? (No. 68, p. 20.) Current system planners are suggesting that core standards or criteria categories include: claims of performance effectiveness; educational significance; evidence of usefulness to others; and indicators of quality that experts (or their surrogates) can observe by becoming familiar with the product, program or practice. In using these criteria to distinguish between promising and exemplary programs, Lois-ellin Datta suggests that:

> *An exemplary program should meet all the criteria for promising and in addition clearly be (among[4]) the best of the population of such programs, after a diligent search for the population. Second, an exemplary program should do more than reliably improve whatever. An exemplary program should make a major, large, educationally block-buster difference in whatever. Improving reading from he 30th to the 39th percentile might be promising. Getting all kids in a class to read at grade level or higher might be exemplary.* (Personal communication, 1/13/96)

- Consumers will need to be represented in the evaluation decisions, but they will welcome judgements of other practitioner, researcher and technical assistance experts . The Findbest System users will be asked to help expert panels decide on priority topics for Findbest searches and to decide on both descriptive comparison categories and evaluative criteria that would be important in helping them make informed decisions. Where feasible, the Findbest System will also develop systematic ways to collect, synthesize and share evaluation information from a range of users.

- In addition to adhering to general principles of good evaluation, consumer-oriented evaluation systems should be designed to make criteria, results, and procedures as fair, understandable, easy to use, timely, visible and public as possible. General evaluation standards are summarized by Govert (1995, Table 4.1, p. 38-39). Since "conflicts of interests" are likely to be a major threat to fairness in the Findbest System, ED's lawyers are researching how other agencies and systems handle this challenge.

The development and application of improved methodologies in consumer-oriented evaluation address OERI's dissemination and evaluation research responsibilities as well as filling a practical need in developing this new system. Some OERI staff and contractors have been involved in various aspects of this research and OERI is commissioning national evaluation leaders such as Michael Scriven and Lois-ellin Datta to help.

---

[4] Lois-ellin, would you be willing to include this modification so that we don't have to find one best, but could find several best with slightly different strengths and weaknesses?

- **OERI needs to clarify how the Findbest System will focus on the evaluation of R&D.**
  Since OERI is the Department of Education's research, development and dissemination office, it should take the lead in learning what R&D is best. This evaluation of R&D is also related to OERI responsibilities for developing standards for the review of its grants and for assessing the performance of ED programs in relation to consumer satisfaction and other criteria. There are two different ways to describe the role of R&D in the Findbest System.

  - First, only R&D-based resources would be reviewed by the expert panels. This means that panels will NOT review someone's great idea, computer software, or a teacher practice for which there is no research or evaluation information to support a claim that it works. Although Expert panels will use research syntheses as they make their decisions, they will NOT judge research findings and other information such as research syntheses and interpretive papers or videos as promising or exemplary unless there is a performance claim. (ED will continue to make these information resources available via the Educational Resources Information Center (ERIC) system, Internet, and other technical assistance mechanisms using existing quality and appropriateness criteria. Exceptions will be made only if there are claims that the use of these research information resources positively impacted performance.

  - Second, in the legislative definition of "programs" to be designated promising or exemplary, "research findings" are included along with products, programs, practices and policies. The Findbest System is proposing to include research findings in this review system indirectly unless they are accompanied by a performance claim as described above. Panels will use research findings indirectly by considering research syntheses as the basis for research based principles which may be used as descriptive information, or when appropriate as evaluation criteria, for the programs to be reviewed. If research principles are included as core or supplemental panel specific criteria they should be quite definitive. For example, inappropriate use of research principles relating to process or style issues can be identified by asking if the desired performance can also be attained using other principles or approaches. [See Datta (Jan. 1996) and Scriven arguments[5] (1993, 1995.]

- **Influencing the norms and expectations for evaluation.**
  A commissioned paper by Gerald Barkdoll on *Standards and their use in the Food and Drug Administration, FDA* (1996) shows how core parts of a federal consumer protection evaluation system can spur the development of supplemental quality control components of the system and that it helped develop a whole industry culture that values and pays for high quality evaluation using controlled clinical trials with randomization, blinding and proper statistical analyses. (p. 8, Barkdoll, 1996).

- **Enhancing education consumers evaluation capacities, skills and opportunities to make wise choices about the resources they use or recommend for use by others.**
  The Findbest System would be a practical way for many involved in education to learn about evaluation criteria and procedures and to practice what they are learning. Unlike teacher or student evaluations, where they are often the subject (evaluand) of the evaluation, Findbest would empower consumers to evaluate the resources they use. More specifically, the Findbest System would:

  - Provide systematic easily accessible up-to-date descriptive and evaluative consumer information on comprehensive sets of promising and exemplary educational resources. The Findbest System would share this consumer information on the promising and exemplary programs via existing personal, print, audio-video and computer-based information providers throughout the National Education Dissemination System. Information on individual programs would be presented using common terms to help consumers compare their options. While this information would be intended to help potential users make informed selection decisions, it would also provide some recognition for the developers and producers of the programs because they would be cited.

  - Help teacher and others' learn how to make wise choices from among the many potentially useful resources as part of their professional development and job responsibilities. The ability to select and evaluate educational resources is included as one of the duties of the teacher in Michael Scriven's model of teacher evaluation (Scriven, 1993?) and in criteria developed by the National Board for

---

[5] Lois-ellin and Michael, what do you consider your most accessible references on this point?

Professional Teaching Standards (check with David Mandel for best reference). Helping teachers learn to find the best was also specified in some unfunded programs that were included in the 1994 reauthorization of OERI such as the Teacher Research Dissemination Demonstration Program and the Goals 2000 Community Partnerships, but instruction in the systematic selection of products and programs could also be incorporated in other federal efforts such as the Eisenhower Professional Development Program.

- Involve education practitioners and other stakeholders including students in contributing evaluative information to the expert panels and subpanels and to nation-wide evaluation feedback systems on the materials they used as described by Campbell (1994). Consumers would also be involved in selecting priority areas for panels to find the best, suggesting what descriptive dimensions and criteria are most important to them and in evaluating the consumer reports and database services provided by the Findbest System. By designing a system for the active involvement of many participants, it is hoped that Findbest will be responsive to consumer needs and help all involved obtain valuable hands-on evaluation experience.

Your Role in Advising on the Findbest System

Readers of this article will have a wide variety of expertise that should help design a system that works for us all. OERI welcomes your suggestions and would be glad to include you in a findbest listserv they have established for authors of the commissioned papers and others who have volunteered to advise us. Please share your reactions with Dr. Susan Klein, OERI, Office of Reform Assistance and Dissemination, 555 New Jersey Ave., NW, Washington, DC 20208-5643 or sklein@inet.ed.gov. If you would like to participate in the findbest listserv, please include a short statement on who you are and why you are interested along with your e-mail and regular address and telephone and fax numbers.

References

Backer, T. E. 1995. *Dissemination Utilization Strategies for Foundations: Adding Value to Grantmaking*, Contributing to the Dialogue: A Paper of the Ewing Marion Kauffman Foundation, Kansas City, Missouri. (Tel. 816-932-1000 or fax 816-932-1100.

Barkdoll, G. January 1996. *Standards and their use in the Food and Drug Administration (FDA)*. Paper commissioned for the OERI Board, U.S. Department of Education, Washington, DC. (Available on WNW, ED.gov)

Campbell, P. B. 1994. *Whose Knowledge is It? Involving Teachers in the Generating and Using of Information on Educational Innovations.* July 10, 1994. U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC. (Also available on ED Gopher. ED.gov)

Canine, D. 1995. Using What We Have to Get the Schools We Need: A Productivity Focus for American Education. A Report from the Consortium on Productivity in the Schools, Institute on Education and the Economy/Teachers College, Columbia University, NY, NY.

Covert, R.W. 1995. A twenty-year veteran's reflections on the guiding principles for evaluators. In *New Directions for Program Evaluation*, Guiding Principles for Evaluators edited by W. Shadish, Dianna Newman, Mary Ann Scheirer & Christopher Wye, No. 66, Summer 1995. San Francisco, Jossey-Bass Publishers.

Datta, L. 1994. *"A Matter of Consensus" in Context*, April 17, 1994. U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC. (Also available on ED Gopher, ED.gov)

Datta, L. 1996. E-mail message to Susan Klein in response to questions from OERI. (Jan. 13, communication from Hawaii to Washington, DC.)

Fournier, D.M. Ed. 1995. Reasoning in Evaluation: Inferential Links and Leaps. *New Directions for Evaluation*, Lois-ellin G. Datta, Editor-in-Chief, 68, Jossey Bass Publishers, San Francisco.

Klein, S. S. 1992 a. A Framework for Redesigning an R&D-Based National Education Dissemination System in the United States. *Knowledge: Creation, Diffusion, Utilization.* (13) 3 p. 256-86.

Klein, S. S. 1992 b. Tilling Fertile Soil. *Knowledge: Creation, Diffusion, Utilization.* (13) 3 p. 330-348.

Klein, S. S. 1993, Ed. Special feature "Sharing the Best: Finding Better Ways for the Federal Government to Use Evaluation to Guide the Dissemination of Promising and

Exemplary Education Solutions" *Evaluation and Program Planning.* Pergamon Press, NY (16) 3 p. 209-278.

Lawler, E.E., 1991. *High-involvement Management*, Jossey-Bass Publishers, San Francisco

Scriven, M. 1993. Hard-won lessons in program evaluation. *New Directions for Program Evaluation.* No. 58, Summer 1993. San Francisco, Jossey-Bass Publishers.

Scriven, M. 1993?. Using the duties-based approach to teacher evaluation. In a volume to be edited by Lawrence Ingvarson, Australian Council on Educational Research.

Scriven, M. 1995. The logic of evaluation and evaluation practice. In Fournier, D.M. Ed. 1995. Reasoning in Evaluation: Inferential Links and Leaps. *New Directions for Evaluation*, Lois-ellin G. Datta, Editor-in-Chief, 68, Jossey Bass Publishers, San Francisco. Pages 49-70.

U.S. Congress. 1994 a. Goals 2000: Educate America Act. Washington, DC March 21. 103 rd Congress, Washington, DC. (Its Title IX contains the reauthorization for the Office of Educational Research and Improvement.)

U.S. Congress. 1994 b. Improving America's Schools Act. 103 rd Congress, Washington, DC. (Major parts reauthorize the Elementary and Secondary Education Act of 1965)

U.S. Department of Education. December 1994. Strategic Plan for the U.S. Department of Education. Working Document, Washington, DC.

U.S. General Accounting Office. 1995. *Schools and Workplaces: An Overview of Successful and Unsuccessful Practices* (GAO/PEMD-95-28, 1995) Washington, D.C.

Weiss, H.B. 1995. *Raising Our Future: Families, Schools and Communities Joining Together*, Harvard Family Research Project

Wohlstetter, P., 1993 *School-Based Management: Strategies for Success,* CPRE Finance Briefs, New Brunswick, NJ)

T

Technical Assistance and the Creation of Educational Knowledge

Brenda J. Turnbull
Policy Studies Associates
March 1996

Under laws enacted in 1994, the U.S. Department of Education is responsible for pulling several new or revamped programs into a mutually reinforcing system of support for educational improvement, incorporating research and development (R&D) along with technical assistance. Traditionally, the federal role in elementary and secondary education has included work in knowledge creation and use--work that capitalizes on the federal government's central vantage point by discovering, testing, and refining research findings and effective practices and helping to build the professional capacity of schools, districts, and states. This paper was commissioned as a small contribution to the cause of strengthening these federal roles, specifically by tightening the working relationship between assistance and R&D. It addresses the question: How might federally supported technical assistance centers form an effective partnership with the educational R&D and evaluation enterprise? The answer I offer, in brief, is that the policy premises underlying both assistance and R&D should change radically to accommodate an entirely different vision of the relationship between knowledge and practice and, correspondingly, between R&D resources and educators.

Federally supported technical assistance in education encompasses a range of responsibilities. A good deal of it is geared to equipping state and local grantees to carry out the requirements of federal programs, especially procedural requirements (Haslam & Turnbull, in press); this work can use some fine tuning, but its relationship to R&D is minimal and not problematic. At the same time, assistance providers have also worked to introduce new ideas about curriculum, instruction, assessment, and organizational arrangements that can bolster student learning. This work does connect in important ways with the R&D realm and is thus of concern in this paper.

The Improving America's Schools Act of 1994 requires ED to establish "a networked system of 15 comprehensive regional assistance centers to provide comprehensive training and technical assistance, related to administration and implementation of programs under this Act, to States, local educational agencies, schools, tribes, community-based organizations, and other recipients of funds under this Act" (Section 13101(a)(1)). The law goes beyond a narrow concept of program administration and implementation: it charges the new centers with developing the capacity of state and local agencies and organizations to define and address issues in educational reform. Rather than focusing their help around categorical programs for special student populations or purposes, the centers are to help schools, districts, and states devise and carry out comprehensive strategies for improvement in teaching and learning so that all students can reach high standards. Among the requirements for the comprehensive centers is that they coordinate services with other ED programs of technical assistance and R&D.

The educational R&D system supported by ED's Office of Educational Research and Improvement (OERI) is also undergoing reorganization in response to new authorizing legislation, the Educational Research, Development, Dissemination, and Improvement Act of 1994. OERI encompasses several National Research Institutes and an Office of Reform Assistance and Dissemination. Its overall portfolio includes funding for university-based centers (usually consortia of researchers at several institutions) that assemble programs of inquiry on teaching, learning, assessment, and governance; regional educational laboratories that offer programs of research, development, and technical assistance for states and localities in their regions; and assorted independent investigations. A premise of R&D policy is that all this work will ultimately pay off, through direct or indirect means, in improved teaching and learning--put simply, that disciplined inquiry produces knowledge that can strengthen the work that educators do. Indeed, part of OERI's statutory mission is "promoting the use and application of research and development to improve practice in the classroom" (Section 912 (a)(2)(3)). The law also offers OERI guidance in the form of definitions (Section 902 (l)):

> The term "development" (A) means the systematic use, adaptation, and transformation of knowledge and understanding gained from research to create alternatives, policies, products, methods, practices, or materials which can contribute to the improvement of educational practice; and (B) includes the design and development of prototypes and the testing of such prototypes for the purposes of establishing their feasibility, reliability, and cost-effectiveness.
>
> The term "dissemination" means the communication and transfer, through the provision of technical assistance and other means, of the results of research and proven practice in forms that are understandable, easily accessible, and usable or adaptable for use in the improvement of educational practice by teachers, administrators, librarians, other practitioners, researchers, policymakers, and the public.
>
> The term "technical assistance" means assistance in identifying, selecting, or designing solutions based on research to address educational problems, planning and design that leads to adapting research knowledge to school practice, training to implement such solutions, and other assistance necessary to encourage adoption or application of research.

As it reorganizes, OERI is trying to reframe its relationship with educators. Public statements from its leadership feature new language about the primacy of customers--defined as teachers, school administrators, and researchers--and are intended to signal an increasingly useful role for the agency. Two-thirds of the members of the new National Educational Research Policy and Priorities Board are teachers or administrators. The board and the agency's staff are seeking to formulate a view of R&D as a usable resource for customers in schools, and to make policy decisions that take sensible account of these customers' capacities and needs.

In my view, a partnership between R&D and educators--which would encompass a partnership between R&D and technical assistance--will only work if it is approached in an unconventional manner; this paper sketches my ideas. I believe that federal policy on R&D and technical assistance has been founded on faulty premises, to the detriment of both enterprises. The conventional design for connecting R&D with practice via technical assistance is a linear one that begins in the academy and proceeds to the field. It pursues an aim of technology transfer--not "technology" in the sense of computers, modems, and the like, but in the sense that new pedagogical skills and organizational arrangements are thought of as technologies that move intact from place to place. We have policies promoting the widespread adoption of effective models of practice, on the assumption that a research-based innovation or a well-evaluated local program can and will be replicated in other schools--that proven practices, by traveling intact to many other sites, offer a lever on educational improvement that is both powerful and efficient.

Much of the conversation about scaling up effective practices and programs reflects this linear model of knowledge creation and use. It assumes that knowledge is largely created through the scientific techniques of research and evaluation--techniques whose most legitimate practitioners are outside schools. The policy discourse further assumes that knowledge derived from research or evaluation has to be translated into a more user-friendly form, featuring practical "how-to" instructions; and that dissemination and technical assistance are the means of shipping packages of knowledge out for widespread implementation.

In keeping with such a model, current technical assistance includes many events that expose practitioners to a smattering of new ideas. When assistance providers have the luxury of spending more than a little bit of time at a site, their efforts are typically aimed at helping with the installation of a new model program or practice. Researchers and developers are, correspondingly, expected to firm up their model programs and practices in prototype testing so that these models can be turned over to dissemination specialists. The Program Effectiveness Panel devotes many person hours to reviewing and "validating" the soundness of the written evidence--assembled at a cost of many more person hours--that programs have achieved their intended aims; only then can a program's developers gain access to National Diffusion Network funding for dissemination and technical assistance.

I will argue here that a lot of this work, no matter how skillfully executed, is wrong-headed. The chief target of my argument is the policy assumption that the R&D community and the evaluators of exemplary practices produce knowledge and educators in the field consume knowledge. Although no one endorses the linear model of knowledge production and use in its starkest, caricatured form, the addition of feedback loops does not make it any less linear or any more helpful as a representation of social learning. In reality, everyone in the system--teachers, administrators, assistance providers, researchers, and policymakers--has a lot to learn about ways of reaching ambitious aims for student performance. To label some of these parties "producers" of knowledge and others "consumers" is to subscribe to antiquated expectations about roles. Instead, we need to expect inquiry and invention to occur everywhere.

The statutory definitions of development, dissemination, and technical assistance are, in fact, broad enough to encompass this more egalitarian and participatory view. What would need to change would be the conscious and unconscious assumptions on the part of program managers who have developed a good deal of program literature (performance indicators, budget justifications, requests for proposals, invitations for commissioned papers like this one, and so on) around an implicit division of labor in knowledge production and consumption.

I believe the old-fashioned assumptions buried in this literature weaken the case for both technical assistance and R&D by portraying them inaccurately. These enterprises are then held to unrealistic standards while some of their genuine accomplishments can go unrecognized. For example, it is not sensible to expect development or validation to generate packaged programs or practices that can be disseminated through a routine, low-skill, low-cost process. Investments in development and evaluation should not be premised on the argument that they yield the proverbial better mousetrap, because then technical assistance has to be portrayed as some combination of infomercials and consultant services, and serious federal investments in it seem superfluous. (If these products are so great, why wouldn't the state and local recipients pay the nominal costs of adopting them?) By the same token, conventional arguments for R&D and technical assistance neglect to portray either the challenges or the productivity of the field-based process of assimilating and refining new ideas. Yet this process, often when assisted by external organizations that bring certain technical skills, is arguably just what can most effectively strengthen our schools, districts, and states. OERI should become better acquainted with this process and, I believe, should celebrate it as a means of learning and invention that deserves adequate federal support.

In this paper, I elaborate on the evidence that invention does in fact occur everywhere and on the practical implications for both assistance and R&D. I also take issue with static notions of educational innovations as products and of producer-consumer relationships. Ironically, these notions are loosely borrowed from the commercial realm, where in fact a lot of new thinking about customers has made these ideas obsolete.

The Evidence of Reinvention

If anything should be clear from nearly three decades of research on the implementation of social programs and model innovations, it is the fragility, instability, and futility of efforts to install whole programs intact in new sites. Practitioners and schools are not blank slates on which researchers and social engineers can draw new designs. They have personal and organizational histories into which they assimilate--gradually, piece by piece--new knowledge that comes their way. This is clear, for example, in the classroom research on the implementation of the California mathematics frameworks, which were designed to embody a far-reaching overhaul of teaching and learning but instead have added incrementally to the thinking and practices of teachers (Cohen, 1990).

A seminal line of research on the implementation process was that of Hall and Loucks (1978), whose Levels of Use framework depicts teachers using classroom innovations in a sequence of partial ways. Some ignore a proffered innovation. Some carry out parts of it mechanically. Later, some use it in a routinized way, with or without incorporation of its most demanding or unconventional elements. Finally, some not only fully incorporate the innovation into their teaching but--significantly--modify it as they reflect on its strengths and weaknesses. An image of faithful replication of a model program is not prominent in the Levels of Use sequence. At most stages, replication is partial, whether for reasons of personal convenience, critical thinking, or some combination of these and other factors.

Finally, let us consider the implications of evaluations of demonstration programs. In education, the granddaddy of these studies is that of the Follow Through Planned Variation program. A mountain of expensive and hotly contested evidence from student testing and implementation studies yielded only a few lasting conclusions, among them this one: there was more variation in effectiveness among the sites supposedly implementing the same model than across models (Anderson, St. Pierre, Proper, & Stebbins, 1978). Thoughtful reviewers of the evidence from planned variation studies and other demonstration programs have concluded that classroom practice is a realm that may not lend itself to the comparative testing of models because of the complexity of the treatments and the disagreements that arise over measuring and weighing their multiple outputs (Rivlin, 1974).

There is an ironic footnote to this argument about the futility of presenting schools and teachers with elaborate prescriptions derived from R&D. Not only are such prescriptions ill suited to implementation by human beings working in organizations (for reasons inherent in both the people and the organizations), but in fact they do an injustice to the research enterprise as well. Social research at its best does not generate tidy packages of immutable conclusions or action steps. Instead, research is the continual testing, refinement, and reformulation of probabilistic propositions. Good researchers seldom cling dogmatically to a simple finding; instead, they are more often drawn to the exploration and understanding of apparent paradoxes and contrary evidence. Sloganeering about research findings and model programs dumbs down research just as it dumbs down teaching.

Despite the evidence that the replication of model programs is an aim that does injustice to both practitioners and researchers, federal policies remain in place to pursue it. Dissemination and "going to scale" remain rallying cries, and some policymakers want to enlist technical assistance in a long-term, labor-intensive process of helping educators carry out tested models more faithfully. The skills of evaluation and expert judgment are enlisted in the trivial cause of determining which pilot or demonstration programs should have their footprints set in concrete for wide-scale replication.

If the idea of replicating well-researched models appeals to policymakers by virtue of its apparent efficiency, this is the efficiency of Henry Ford, and it is ill suited to the times we actually live in. Let me suggest that a different set of ideas would more productively organize the relationship between R&D and technical assistance. I think that some of the more intriguing ideas come from the business literature, and it is to those that I now turn.

Post-Industrial Lessons from Business

A model program emerging from several years of R&D or evaluation, with one-size-fits-all instructions intended to cover every practical contingency, is a product like a Model T, and assistance providers who tell educators about it and help them install it are providing a service in a way that would have been familiar to the advertisers and retailers of the early twentieth century. Business has moved on from those days, however. In particular, the companies that are innovators in marketing and R&D have reorganized their relationships with customers in ways that could have parallels in education. In drawing these parallels, I am thinking of the federal government (not the individual organization providing assistance or R&D) as the analogue to the company: the government's R&D and assistance programs, like R&D and marketing divisions of a company, are collectively responsible for offering something of value to customers. The ultimate aim is different--the company seeks a profit by displacing its competitors' products or services, while the government seeks more effective teaching and learning by displacing the status quo--but in both cases an important intermediate outcome is widespread use of the company's or government's offerings, and in both cases success depends on leadership that enables the whole enterprise to make an effective connection with customers.

One concept worth considering is that of involving the customer as a participant in the creation of value. Traditionally, in business schools as in the world of educational R&D, a linear notion prevailed. Companies were said to create things of value, which they handed off intact to relatively passive customers. As anyone who has shopped at IKEA or withdrawn cash from an automated teller machine (ATM) knows, the boundary between the company and the customer has become more permeable. IKEA, a furniture retailer whose huge, no-frills stores sell products for do-it-yourself assembly, meets a need by recognizing that customers do not necessarily want to pay for intensive sales help or furniture assembling; they are capable of selecting their own groupings of furniture and putting the pieces together, and many will choose to do business with a company that assigns them these roles while offering lower prices and amenities like child care and a cafeteria in the store.

As Normann and Ramirez put it, "IKEA wants its customers to understand that their role is not to *consume* value but to *create* it" (1993, p. 67). Normann and Ramirez also cite ATMs as an innovation that allows the customer to participate in creating a new kind of value--namely, round-the-clock access to cash. Curmudgeons who dislike assembling their own chairs or facing cashless machines on weekends may disagree that IKEA and ATMs provide a net increase in value, but the fact remains that today's customer has assumed a larger role in creating value, and this role is premised on business's recognition of the capabilities of the customer. For successful companies today:

> Their key strategic task is the *reconfiguration* of roles and relationships among this constellation of actors [suppliers, business partners, allies, and customers] in order to mobilize the creation of value in new forms and by new players. And their underlying strategic goal is to create an ever-improving fit between competencies and customers. (p. 66)

The federal government in its support for educational R&D and assistance has not yet addressed the reconfiguration of roles, even though such a reconfiguration is overdue because the customer for educational R&D or assistance is already, inescapably, a key creator of value. Not every purchaser of furniture wants or needs to assemble it all herself, but every principal or teacher really does have to assemble new skills into a flow of accomplished practice. This suggests that policymakers should discard the assumptions about value creation that are implicit in a linear model of R&D and assistance; instead, they have to recognize that customers are also creators (not primarily consumers) of value.

A second concept to note in the business literature is the erosion of the distinction between products and services. IBM, which used to see itself as being in the business of developing and selling products, increasingly positions itself as a service company: "the customer can buy any box from any vendor and IBM will supply the systems know-how to make the whole thing work together" (McKenna, 1991). Similarly, policymakers should consider whether the results of educational R&D are best understood as being analogous to products, or whether researchers, developers, evaluators, and technical assistance providers might not be most productive when they envision themselves as service providers. Field-based development, as I will illustrate below, is a recursive process of assistance, development, and theory building. If we were to recognize this in articulating R&D policy, the implications for ideas about customer relations would be profound: rather than relatively passive makers of choices (i.e., adopters of products), customers would be understood as actors in the more ongoing, reciprocal relationship that characterizes a successful service business.

Third, and related to both of these concepts, is the transformation of marketing and of the relationship between marketing and product development--especially in the more innovative high-technology companies. The pace of product development in such companies is astonishingly fast. The advances in technology that permit far greater and more rapid customization of products enable them to reach more customers, but along with product diversification comes an overhaul of traditional marketing. Companies realize that they must not only cultivate service relationships with their customers but, more radically, deploy their sales representatives as part of the R&D team so that they understand their customers' situations in detail and new products are jointly created by the company and the customer (McKenna, 1991). Rather than just allowing an R&D department to engineer new variations on existing product lines, or just surveying customers about their needs, some corporations engage customers more interactively in the design and refinement of new products--and, significantly, place their own technical experts in situations where they will learn about the daily realities that customers face (Brown, 1991; Hamel & Prahalad, 1991).

Similarly, R&D and technical assistance in education can most effectively meet their customers' purposes if they join forces to continually fine-tune products and services in close consultation with the customers themselves rather than hoping to present the field with a few products perfected through impeccable research designs. In such a process, developers would learn what they need to know about customers' environments, routines, and predispositions; customers would be in a position to help create value (i.e., to contribute to a more decentralized process of knowledge building) through their own learning. In words that could be used to describe good technical assistance (and, for that matter, R&D) in education, McKenna describes the new form of marketing: "it is based on developmental education, incremental improvement, and ongoing process" (1991, p. 66). It will be ironic indeed if policymakers in education fail to recognize that learning--on the part of developers and customers--is a long-term and interactive process not to be dismissed as a simple form of dissemination.

In education, if we can imagine customers--educators--as more active participants in development, we accept the reinvention of practices in new settings as a source of creative adaptation rather than as a minimally necessary appendage to development or, worse, as the enemy of quality control. Marketing (or technical assistance) is then a process deserving respect as the bridging of the boundaries that isolate educators on one side and knowledge creators on the other; all are engaged in finding solutions to problems of practice. If education R&D and technical assistance are aimed at scaling up the installation of a limited set of model programs--limited because of their ponderous development cycles and the rarity of rigorous evaluation--they are missing the best opportunities to advance their customers' learning.

Some Examples from Educational R&D

Although the policy discourse about R&D and technical assistance is mired in old, linear assumptions, many practitioners of R&D and technical assistance are actually engaged in day-to-day activities that fit more appropriately into a late-twentieth-century model of the relationship between service providers and customers. My examples are drawn from the work of regional educational laboratories, which Policy Studies Associates recently evaluated (Turnbull, McCollum, Haslam, & Colopy, 1994). In this evaluation we analyzed examples of educational development--sometimes the forgotten half of the R&D acronym. Here, I emphasize our findings about (1) the relationships that developers formed with participants in the field and (2) some different ways of viewing knowledge creation.

Most of the examples of development that we studied in the laboratory program were striking in the depth of their partnership with field sites. Participants at the field sites had a role that went far beyond the traditional one of experimental subject, and they praised the laboratories for recognizing practitioners' capacity to contribute, as professionals, to the development process. The comment of a participant in Integrating Education and Human Services (a development effort of the Northwest Regional Educational Laboratory) captures this response nicely:

> When we administered the first draft of the survey for the community profile, many people found it to be very difficult--too long, too wordy, not applicable if the person was not from a service provider agency. I approached [the project leader] reluctantly, telling her that there were some problems. She said, "No problem, tell me about it. That's your role; you are a pilot site." She was very good about using the feedback, and also very good at providing feedback to our members on the analysis from the self-study profile. She did it without making people feel bad. It was very constructive.

Similarly, a staff member at one of the pilot sites for the Mentoring Guidebook of the Regional Laboratory for Educational Improvement of the Northeast and Islands said: "The process was interactive from the beginning.... [Laboratory staff members] had great conferencing skills and really knew how to ask the right kind of open-ended questions. In other words, they modeled the mentoring process for us." This development effort called on state staff to become familiar with research literature and synthesize it into a new written product (with conceptual and editorial help from laboratory staff). In later stages, the local pilot sites initiated many revisions to the written materials. Moreover, their interaction with the states and with one another proved so valuable that the laboratory reframed the overall effort to include support for a continuing network among practitioners.

A principal at a pilot site for SouthEastern Regional Vision for Education's Formative

Teacher Evaluation said:

> Most [outside consultants] want you to accept their program, but [the laboratory staff] allowed us so much freedom ... but we would still get help and support at a critical time.... They respect practitioners' opinions and views. That attitude is not always present at the university level--they'll tolerate you.

Laboratory staff worked closely with this principal, his colleagues, and educators in two other districts during the pilot phase, presenting workshops and following up with individual help. The pilot districts and schools, in turn, evaluated their implementation of the new techniques--using evaluation designs that they created themselves, with help from the laboratory--and reviewed drafts of the handbook. One- or two-day workshops that introduce new districts to the idea of Formative Teacher Evaluation were subsequently conducted by central staff and teachers from participating districts; laboratory staff helped organize this dissemination effort but shared the lead responsibility with the field-based practitioners.

A school district staff developer who has led case discussions as part of the Case Methods work of the Far West Laboratory for Educational Research and Development told us, "I've often heard [a laboratory staff member] say how much she has learned from us"; this participant also called the laboratory "a great resource." According to numerous participants in the process, including practitioners as well as laboratory staff members, frequent reflection and interaction on the successes and failures experienced in case discussions have enabled all parties to contribute to the revision of draft materials and also to more clarification and elaboration of the advice available to future discussion leaders.

As the quotations from active participants suggest, the mutually respectful interactions between participants and laboratory staff contributed to an interesting phenomenon: by involving field sites in the development process, laboratories could very effectively deliver technical assistance to those sites. Each of the individuals quoted above (as well as others whom we interviewed in the evaluation) made this connection more or less explicitly. The participant in Integrating Education and Human Services pointed to an exchange of feedback on an even footing, where site personnel gave the laboratory a critique of the draft materials, and then the laboratory constructively summarized the analysis of the site's profile. The participant in developing the Mentoring Guidebook said the laboratory staff behaved like skilled mentors. The principal who pilot tested Formative Teacher Evaluation praised the laboratory for simultaneously conveying respect and providing help and support. The staff developer involved with case discussions recognized that the laboratory staff learned from the field at the same time as they were a resource.

In these sites and others we visited, a relationship of reciprocal assistance was a somewhat surprising experience for participants, but they spoke highly of such a relationship as the context for receiving assistance. In the Partner Schools program of the Northeast Laboratory, where a reciprocal relationship was a cornerstone of the activity's design, many participants responded positively to it. As one said:

> The lab has great ideas, but this project is strictly give and take. There's real interaction. The lab always wants to know what worked for you. They're drawing best practice ideas from us–this is not a one-way project.

These findings from the laboratory evaluation indicate that some members of the R&D community, at least, are actively involving their customers in the creation of value. Moreover, the nature of the development work that we studied in our evaluation leads to some unconventional observations about the nature of knowledge creation in field-based R&D. The laboratories' activities were strikingly eclectic with respect to both the kinds of practical and theoretical tests applied in field testing and the amount of local variation expected to persist during implementation. Although the very largest and longest-term R&D efforts embody formal procedures for knowledge building, a host of smaller ones also rely on the improvisational skill of R&D professionals who approach their work with a combination of theory, practical savvy, and flexibility. Three examples of small and medium-sized efforts can illustrate some actual knowledge-development processes found in laboratory work.

Study Groups organized by the Appalachia Educational Laboratory reflect the refinement, through field experience, of a set of standard operating procedures by which participants (usually teachers, in the Study Groups that we examined) can become acquainted with a body of research in some depth. In the early years of Study Groups, laboratory staff members learned several practical lessons, which they codified in a 1987 report: for example, that groups should be no larger than about ten participants, that they should develop their own processes and products, that costs should be shared, and that laboratory staff should facilitate each group's organization and functioning. By the 1990s, the procedures for Study Groups were well established and not subject to much further refinement, although each group provided evaluative feedback on its experience.

Formative Teacher Evaluation looked quite different as a development effort. It was a process in which field experience helped to give more concrete form to "a really open-ended" idea, as a project staff member described it. Laboratory staff emphasized with some pride: "We don't have a canned Formative Teacher Evaluation plan. We provide some options and some methods, and they [in the adopting district] develop a plan." The handbook developed by the team, with considerable input from the first three districts using Formative Teacher Evaluation, described the operational specifics of the plans developed and carried out by these districts; it was designed to include the theory behind the process but to emphasize a practical experience base. Nevertheless, the concrete shape given to Formative Teacher Evaluation in the handbook was not the end of the development road, according to laboratory staff. They expected a further evolution in these and other districts. And, unlike some other laboratory products, this one was apparently expected to remain open-ended forever; there was no press to use field experience as a way of identifying more effective procedures and then to recommend or mandate their use. In this activity, then, development was almost synonymous with local implementation, which was expected to reflect unique local priorities and preferences.

The developers of Case Methods are contributors to the published literature on teacher preparation and professional development, and their field experience with the writing and discussion of cases reflected yet another approach, one with more of a flavor of applied research. These developers--like some of their colleagues in laboratories nationwide but unlike others--are part of an ongoing conversation with researchers in universities and other laboratories. Research on professional development, such as the literature on organizational and individual change that supports critical and collaborative cultures among teachers, gave these developers criteria by which to assess their accomplishments and refine their methods. A theoretical framework thus helped them make sense of the results found in field testing and guided the revisions of the Case Methods products and processes.

Across these three activities, then, we see different concepts of knowledge creation. The development of Study Groups culminated in a relatively fixed set of standard operating procedures--yet, at another level, we can characterize these procedures as the vehicle that enables future groups of participants to create their own knowledge. The development of Formative Teacher Evaluation suggests a view that systematic knowledge creation ends with the launching of an open-ended idea that permits others to make local adaptations; the laboratory staff expected each site to adapt the basic design extensively. (Regrettably, in my view, they described no plans to capture and build systematically on what might be learned across sites.) Case Methods potentially reflect a more active, recursive development process in which the laboratory orchestrated theory and field experience into ever-stronger products and services (although realizing this potential would require a larger investment than the laboratory was making at the time we studied this program).

Despite this diversity, what these examples of development have in common is that almost none of them is best understood in the conventional, linear terms that pop up in policy discussions: developers spend years engineering a product and collecting customer feedback that helps them refine that product; marketing/technical assistance then takes the product to scale in a large number of sites. Instead, we found an intertwining of development and assistance, with all participants learning together, that more closely resembles the R&D and marketing practices being touted in the business literature. Knowledge creation was field based. From this empirical base, I observe that practice is a few steps ahead of policy in the design of working relationships that bridge R&D and technical assistance.

Implications for Technical Assistance

After (1) arguing that there are well-known flaws in the idea of organizing development and technical assistance around the aim of program replication, (2) identifying some alternative organizing principles found in the business literature, and (3) reporting on field experience in education that supports these alternatives, I finally arrive at addressing the formal charge for this paper: to suggest how technical assistance centers might work in partnership with the R&D and evaluation community.

Technical assistance is the bridging of boundaries. Like the marketing departments of high-technology corporations, assistance providers are in a position to work with customers as they participate in creating value--to bring the customers inside the "company" as participants in R&D or, in other words, to involve educational practitioners as fully legitimate *members* of the educational R&D community rather than its targets.

In the realm of educational improvement, what practitioners do to create value is what practitioners do all the time: they run classrooms and schools where children learn--and, if we are lucky, they continually get better at it by learning themselves. Because real school improvement requires teachers and other educators to learn, assistance providers should do things that help learning take place: create--and help people use--opportunities for problem solving, professional conversation, and reflection. The resources they must bring are not primarily products (i.e., fully engineered programs that experts have certified as ready for replication) but services.

This assistance agenda would go well beyond the current policy expectations that telling and showing are adequate means of providing help, and that what should be told and shown is knowledge that was created somewhere else. Federally funded assistance providers now meet these expectations by transmitting information about classroom practices such as cooperative learning and new techniques in assessment. Teachers and administrators like getting the information; they appreciate being kept up to date on professional trends. However, the effects on their work are rarely more than marginal, and the assistance providers know it. (I would not actually argue for abandoning the simpler forms of information dissemination and assistance, which do render a service that customers like; I would just not pretend that these activities offer leverage on school improvement.)

To a much greater extent than is now the case, I believe technical assistance should be encouraged to promote and assist forms of knowledge development that are directly grounded in the real problems and accomplishments of classrooms, schools, districts, and states. For one thing, technical assistance should be seen as a vehicle for a two-way exchange between those whose primary work is research, development, or evaluation and those whose primary work is in schools. What the assistance providers would carry to the R&D community would not be a report of needs in the field but rather a combination of needs, capacities, and insights informally culled from practitioners. Systematizing and formalizing these insights would be largely the job of R&D organizations, although both assistance providers and practitioners should contribute. We observed, for example, that laboratory staff who are contributing members of a research community can convey the findings of their field trials to other researchers (as well as developers) who need to learn from them. The fact that laboratories span the boundary between research and practice does not just equip them to communicate practitioners' needs to researchers; it also enables them to bring an important, applied perspective to theoretical discussions. Systematizing this flow of information would be useful. In particular, it should be divorced from old images of needs assessment, which emphasize the deficits of practitioners rather than their purposes and capacities.

Looking in the other direction, a deeper two-way flow of information would imply changes in the way R&D is portrayed to potential customers. The nature of the assistance relationship reported by participants in several development efforts by laboratories--in which the exchange of ideas and opinions took place on an equal and mutually respectful basis--offers an image of service provision that could be more widely applied in federal technical-assistance programs and in federally funded R&D. It is rare for assistance providers to present educators with findings or model practices that are portrayed as works in progress rather than hard and fast findings, but in fact such a presentation would be more true to the realities of research.

I do not claim that educators would automatically welcome the presentation of R&D as a work in progress. Like many policymakers, educators who have had little or no experience with intensive, interactive technical assistance are likely to respond with initial skepticism at best. They have been taught to expect research-based information to arrive in the form of packages and slogans. Taking in a superficial presentation of such packages or slogans is fairly painless (though generally fruitless), and it is a ritual to which educators are accustomed. Many or even most would initially resent the idea that there is no free lunch at the R&D cafe. However, a genuine improvement effort does require a great deal of work, and we cannot ethically portray R&D as a drastic shortcut.

What R&D and technical assistance offer instead, in my view, is a set of tools--often embodied in a person with a strong background in a research field, in evaluation methods, in process facilitation, or in other specialized skills that can contribute to local purposes--to help guide the local hard work in productive directions. Thus, the value added by technical assistance when it is well integrated with R&D can be seen as help for practitioners in distinguishing between more and less productive change efforts. Clearly, there are differences; just because a school, district, or state agency is changing does not mean it is getting better. Robust research findings, including those that have been applied and refined in formal development, do offer guidance in setting productive directions for change. Skilled implementation assistance brings perspectives that may be lacking in an organization, and can be used to enhance that organization's own problem-solving capacity in lasting ways. Moreover, to the extent that we move to a greater policy emphasis on interactive, field-based assistance and R&D, it will be crucially important to apply and refine appropriate field-based and participatory forms of quality assurance. Process is expensive, and when it is unstructured or not examined critically it can even do harm to participants and the children they serve.

A related idea is to use technical assistance as a means of scaling up something other than the installation of fully researched and developed practices–namely, the process of inquiry itself. This means that technical assistance would contribute to knowledge development through the methods of continuous process improvement or action research, in which inquiry and the use of data are built into organizational routines. Participants in assistance would expect to use research findings or practices that worked in other sites as starting points for inquiry. Networks of practitioners could join in the collection of reasonably uniform data about what they are doing and how it is working out. I am not suggesting intensive or burdensome forms of inquiry for all participants; data-collection instruments such as teacher logs can be refined so that they take only a few minutes for those who prefer to spend no more than that. Participation in the data analysis could be available to all who muster the time and inclination to join in (but would not be inflicted on those who do not). On a large scale, this could add up to a powerful means for communities of practitioners, researchers, and evaluators to test and refine knowledge. This would mean, again, that the teams providing assistance should have expertise in methods of inquiry and a solid acquaintance with the research fields most relevant to their customers' issues.

These techniques of two-way communication and collaborative inquiry are not brand new; they are already found in the repertoires of organizations that are part of the R&D community, including laboratories, the National Diffusion Network, and some boundary-spanning efforts between universities and schools. What would be different would be their full legitimation in the policy talk about R&D, evaluation, and technical assistance. Measures of the adequacy of ED's customer service would then expand to include measures of customers' successes in implementation and perhaps also their engagement in two-way communication or inquiry, not just the alacrity with which assistance providers furnish answers to questions.

Technical assistance, in summary, should aim to do much more than expedite the communication of research findings and model practices that emerge from an R&D pipeline. It should reflect a thoughtful effort to bring customers into the R&D enterprise as active participants. In business terms, we could call actively participating customers "creators of value"; in more familiar education terms, we would call them learners. A strategic initiative to redesign technical assistance around a more realistic understanding of the learning process would produce a very different kind of engagement between knowledge creation and schools. This strategic initiative would result in very different policy arguments for the investment in R&D

and technical assistance and, more important, it could contribute to communication, reflection, and discovery throughout the education system.

References

Anderson, R. B., St. Pierre, R. G., Proper, E. C., & Stebbins, L. B. (1978, May). Pardon us, but what was that question again? A response to the critique of the Follow Through Evaluation. *Harvard Educational Review*, 48, 161-170.

Brown, J. S. (1991, January-February). Research that reinvents the corporation. *Harvard Business Review*, 102-111.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12 (3), 311-330.

Hall, G. E., & Loucks, S. (1978). Teacher concerns as a basis for facilitating and personalizing staff development. *Teachers College Record*, 80 (1): 36-53.

Hamel, G., & Prahalad, C. K. (1991, July-August). Corporate imagination and expeditionary marketing. *Harvard Business Review*, 81-92.

Haslam, M. B., & Turnbull, B. J. (in press). Issues in federal technical assistance policy. *Education and Urban Society*.

McKenna, R. (1991, January-February). Marketing is everything. *Harvard Business Review*, 65-79.

Normann, & Ramirez, R. (1993, July-August). From value chain to value constellation: Designing interactive strategy. *Harvard Business Review*, 65-77.

Rivlin, A. M. (1974). "Allocating resources for policy research: How can experiments be more useful?" *American Economic Review*, 64, 346-354.

Turnbull, B. J., McCollum, H., Haslam, M. B., & Colopy, K. (1994) Regional educational laboratories: Some key accomplishments and limitations in the program's work. Washington, DC: Policy Studies Associates.

U

May 17, 1996

REPORT on a SPECIAL Office of Educational Research and Improvement (OERI), AERA PRESESSION

by

Allen Schmieder, ORAD, OERI

"OERI Leadership Efforts in Designing Approaches to Identify and Share Promising and Exemplary Products, Programs and Practices"

Monday, April 8, 9:30-11:30 AM, New York Hilton, New York, New York, Presession to the 1996 Annual Convention of the American Educational Research Association

Sharon P. Robinson, Assistant Secretary, OERI, U.S. Department of Education, Chair

WELCOME, INTRODUCTIONS, CHARGE TO THE GROUP

Sharon Robinson, Assistant Secretary, OERI. Good morning. This pre-AERA workshop gives us an opportunity to interact about one of the most challenging current opportunities in the Office of Educational Research and Improvement (OERI): the opportunity to develop standards for the field--broadly, well articulated, well founded, meaningful standards for selecting quality educational programs. Congress has given us the responsibility of identifying, evaluating, and widely sharing the best educational materials and programs in the nation--both from the public and private sectors--so we felt it was very important to gain input from some of the nation's leading experts in this area, who we knew would be attending AERA--to have them react to our intent and proposed procedures and make recommendations on how we can improve our efforts.

Our plan for this morning's program is to provide you with a comprehensive briefing on our progress to date, then have you form smaller groups around particular themes to interact about what we are proposing and identify questions for any of the speakers. After your discussions we will again form a single group so we can answer your questions.

We will start with Congressman, Major Owens, who was the chief architect of the new OERI legislation which lays the foundation for new system we will be discussing this morning. Major Owens will be followed by Alba Ortiz who is chair of the OERI Board Standards Committee, which has been advising us on our proposed OERI Standards to Designate Promising and Exemplary Programs. Bob Stonehill from OERI will generally describe the new National Educational Dissemination System (NEDS); Eve Bither will describe our developmental process to date and introduce you to some of the ideas that have been presented for how "our expert panels" will work and how we will communicate about products and services that survive the evaluation process. Sue Klein will talk about some guiding principles that have been developed in papers we have commissioned and from other activities in OERI. We will have summary presentations on the two pilot expert panels that we are forming. Pat Campbell will cover the Gender Equity Expert Panel and Beatrice Berman and Senta Raizen will go over the developments of the expert panel on mathematics and science. Following these brief presentations, we will ask you to form groups to discuss the proposed plans and to develop questions and recommendations for us around the meaning of this new system for 1) teachers in the field; 2) developers/publishers; 3) for R & D technical assistance providers; and 4) for researchers and evaluators.

It is now my pleasure to introduce Congressman Major Owens who has been a wonderful and very positive force in leading us all in our collective efforts to reinvent OERI. He will give special focus in his remarks to NEDS and relationship to Findbest (the evaluation component of NEDS). He will provide a national political overview and summarize the Congressional intent for this important new challenge.

Major Owens, Congressman, 11th District, Brooklyn, New York, Major Architect of OERI 1994 Reauthorization Legislation. Thanks for inviting me. I always come to this kind of meeting with great reverence and humility. I don't pretend to come to this AERA gathering as a scholar and warn this wonderful audience that my remarks will be primarily political, political in the service of educational research and development. During my time in Congress, especially as chair of the Select Committee on Education and Labor, my best utilization of politics and the best service of politics has been uniting both parties in an effort that has overcome partisan political obstacles to help strengthen support for educational research. Even under the Bush Administration, we were able to get the Republican leadership on the Education and Labor Committee to support an initiative to restructure OERI and not much has changed after the new administration came in. We already had achieved bipartisan support for our new plans for OERI. There have been some changes since the Clinton Administration took over, and, incidentally, not all for the better.

I only mention this to underscore how solid the bi-partisan support has been for our efforts, when we had some differences within our own party, the bi-partisan support within the committee helped us overcome those problems. The support of the scholarly community also helped. People said we couldn't do what we were trying to do. The last thing the power groups wanted was more input from a policy group representing a cross section of the field. But we achieved our goals. Reason won over petty politics.

The present problem that still hurts the Department of Education is that they still have no appropriation. The Department has been singled out by the Neanderthals who wanted to eliminate the Department. They have put it in a difficult position by using the appropriations route to throw up obstacles. They couldn't accomplish their negative aims through the Reauthorization agenda "to remake America." I will make some partisan statements on these matters this morning. But you will have many speakers during this convention so there will be some balance in the views that you hear.

In American education today, there is no need for sounding the emergency alarm, there is no need for a crisis strategy. We need evolutionary changes and improvement. We don't need the kind of revolutionary actions that have been proposed by some, like getting rid of the Department of Education. These kinds of proposals are damaging. There are many days when the Department employees were labeled as nonessential. I want to congratulate all Education Department employees for keeping a steady focus through these difficult times. It is also important to note that even though the majority party in Congress doesn't like the Department of Education, the good news is that no one wants to assault OERI. They have attacked many other programs and units in the Department but not OERI. It is a credit to the new structure that we have put in place. Even when the opposition had drawn up blueprints for the dismantling of the agency, they had places they wanted to put and maintain OERI.

Our bi-partisan group, in arguing hard and long with the opposition, has emphasized that there are a lot of pieces in place that make a lot of sense, ERIC, the Educational Laboratories, the National Research Centers. These quality programs and places have added to OERI's impregnability. These and so many other good educational initiatives supported by OERI have added to our defense against the Neanderthals. They have essentially given up and taken their attacks elsewhere.

The common sense of the American people has prevailed over the last several months. They didn't like the idea of a partisan group shutting down the government in order to "get their way." The President has held the high ground. He wants an increase in spending and support for education. In his State of the Union, he called for the rewiring of schools, 20% of them by the end of June, all 110,000 of them by the year 2,000. Some agreements supporting his stand have occurred since his speech, but cuts are still planned for Head Start, Title I, and other programs. Gains have been small, a trade off with America Corps, Goals 2000 was put back in, etc. Many of us feel, however, that we could have gotten an even better deal. There are still too many cuts dropping into place. Title I and other Elementary and Secondary Programs are still in danger of being diminished. I hope the White House will even further stiffen their resolve.

Money is not the issue. We've just learned that CIA has $2 billion it didn't know it had. I suggested that the two billion go to the Department of Education. Some colleagues still don't know where we will find the needed money. CIA probably has more than two billion it doesn't need. Maybe a lot in petty cash! They fired two people for helping reveal the surplus. The Federal Reserve has $2.7 billion in "reserve." Greenspan is liked by both parties. So there is evidence that we have the money if we need it. The Republicans will continue to play "bad cop, good cop," pushing their harsh education agenda in the House -- and the Senate will be compassionate and agreeable. Despite this posturing the President can and will take credit for major successes in retaining a quality Department of Education budget.

This battle of extremes is probably over, so I think you can expect an Education Budget to be in place in a few weeks, not too long after Congress reconvenes. (Note, this happened on April 25, 1996 with support for Goals 2000, but elimination of funding for equity programs and activities such as the Women's Educational Equity Act and the State Civil Rights Act Title IV Equity Coordinators.)

Now to the educational substance of what we are here today to discuss. The distribution and dissemination of good educational information and the need to have a judgement made about what it is we disseminate. The need to share exemplary products, programs, and processes. There is nothing more important in these days of soft information. I am a librarian and so have a special perspective on how people are being overcome by trivia. We need someone to sort the avalanche of new information. We already have ERIC but it, by design picks up a broad array of educational materials without making differential judgements about their quality. The day has arrived when we have to find a way to identify and share the best and most essential educational information.

The nation's attention has been focused on the educational summit which has included Governors and CEO's of major corporations. Their situation presents a good lesson for education. For a long time now these leaders have not been interested in "everything." They are only interested in what they need to know to make good decisions and provide good leadership. Somebody has to make judgements about what to provide them on a day to day basis. Someone has to sort and refine large amounts of information and make decisions about what is exemplary and promising--on what materials the CEO will most need to make sound decisions. Someone or a group of someones make it possible for people on the firing line to search the data bases and find the information that is most essential for decision makers to see. They will boil down large amounts of information to two or three choices.

Back home in Brooklyn, I was finally able to connect my home televisions to cable. I suddenly had 70 channels to choose from. I started frenetically flipping channels and most of it was "junk." I was overwhelmed with trivia. It is the same in education. We need to get on with the process. We need to start now to experiment on Education 2005 on a mass basis. We need to narrow our choices to those that are most promising and use a process of experimentation to test them. What works in Oregon may not work in New York. We need to test it in New York and in Atlanta.

We also need some kind of certification process. To designate something as exemplary means that it is worthy of reproduction. Some programs may be at an embryonic stage and need funding help--or they won't grow into a full baby. So we will need incentives so people will want to report rapidly, have their program put on a platform and examined. Most won't want to do this unless there is some incentive, some process to move programs from promising to exemplary and then to "institutionalization." Good programs won't be lost, they will be "institutionalized," and make a sustained and sometime broad impact on the quality of American education.

I will close with one example. I am very impressed with the experimental work of Dr. Comer (Comer School Development Program). He is doing an excellent job of dealing with whole schools. We also need to deal with whole districts. Exceptional schools are a high concern of Congressman Hawkins and those working to improve the quality of education for African Americans. There is much effort here, and it is good, but we also need to find the exceptional school districts and better share their successful programs. Focusing on a single school has a limited impact.

Talking to Dr. Comer about his program model brought to mind another example of what we are trying to accomplish with this new legislation. In one of his experimental schools, a room was set aside as a "quiet room." It was used to bring together at risk kids and academically successful and other kids who work together to ensure success of all students involved. It was a simple procedure. And it worked. Have we saved it? What happened to it? It would be possible to do elsewhere. Has it been replicated elsewhere? This simple practice and even more complex and costly programs need to be replicated.

In New York we have tried every single conceivable experimental process that might lead to educational improvement--but most have been lost with time. We have no system for stamping programs as exemplary and promising, no certification procedure. So most of what we have learned and could have used to improve our schools has been lost.

So while high technology is on the minds of our leaders--the educational industrial complex is happening--nothing will be lost, because when all those people begin to wire all those schools the President wants wired, they will find out that the schools are falling down and that the teachers are being overwhelmed with information. You and your colleagues are badly needed. Your expertise will be vital to an educational-industrial complex. You will make judgements about the practical utility of quality instructional materials for teachers who want to help the children of America!

Alba Ortiz, Chair of OERI Board Standards Committee. Good morning. I am happy to tell you that I can condense my remarks because Major Owens did such a great job of telling you why we need standards for identifying and sharing quality educational products. The OERI Board gives the public a voice in national educational research. The Board has 15 members, 5 are well known educational researchers, 5 are from prominent professional schools in education and 5 are selected because of their broad-based experience in education. We have a number of important tasks that include: 1) assessing the state of knowledge in educational research and we will disseminate a product summarizing the state of the art and laying out priorities for assessment and dissemination; 2) finding ways to link research and practice, especially regarding how to best connect consumers with vastly increasing educational resources; and 3) giving guidance as to what constitutes good practice and products.

We are working with Assistant Secretary Robinson to develop standards so that we will have assurance that activities and products recommended meet the highest standards of professional excellence. The Board has already developed standards for processing and evaluating grant, contract, and cooperative agreement applications to OERI. We are in the process of developing, and you are helping here today, the process and criteria for designating programs as promising and exemplary. We will then develop a third set of standards that will be used to judge the performance of recipients of grants, contracts and cooperative agreements.

We want to provide assurance to the public that exemplary and promising really meet rigorous criteria that will allow educators and the public to use the designated materials and programs with confidence. The Board's role is to review and approve the draft standards before they are released for public comment. You each have a copy of the draft standards that we approved at our last meeting.

The document that you have specifies the process that we will use. The law defines programs very broadly, so the kinds of materials to be considered will include such areas as policy on educational research. The process calls for a standing group of experts that will be at the center of a much broader network of expertise about education. We will draw panels from this network that relate to particular programs and activities that are submitted and they will review them and make a judgement relative to their overall success, educational significance and usefulness to others, and how they might be adopted or adapted from one place to another. We are particularly interested in your comments on the criteria we have recommended for distinguishing between promising and exemplary programs. Are there any additional criteria that we should add, or any that we should eliminate? We would also like your opinion of our definition of an educational "program." We will take your and the publics' comments very seriously in developing the final standards. We want the draft disseminated as widely as possible. It will be published in the Federal Register by the end of the month. After all of the comments are in, the Secretary will make judgements about them, include those that improve the process and list and give explanations for those not incorporated. The Board will review the final recommendations regarding comments from the field and approve the final regulations. Once they are signed by the Secretary, they will become binding and will then be used by the Department of Education to designate programs as promising or exemplary. So the public comment period is important. I was involved in the development of the standards for applications to the Department and was surprised that we received only five or six comments. So I look forward to today's discussions as we continue to forge final standards for judging successful educational practices and programs.

Eve Bither, Acting Director, Office of Reform Assistance and Dissemination (ORAD), OERI. In our work in ORAD and OERI we are asked many questions by the public and by Congress. One of the most common, is: "What has research told us about the most pressing problems in education, and what can you tell us about what might work in regard to (and you insert your own problem, need, or challenge, e.g., what is best in elementary science, in reading, what are some of the best teacher inservice education programs, what are some programs for raising the self esteem of middle school students, and on and on ad infinitum). We need standards that will give a quality assurance to programs that the Department of Education might recommend in response to those questions.

You have a copy of the draft standards and I would like to make a few comments about the document.
--How are we planning to use the plan?
--What criteria will we apply?
--What is the difference between promising and exemplary?

I especially liked Major Owen's characterization of the challenge, that he so eloquently summarized at the end of his presentation. I will use it in all of my future presentations!

Now to what we have done so far. As Alba Ortiz related, we will use highly expert panels to determine exemplary and promising programs and practices. These experts will be nationally recognized in research, school governance and school practice. The criteria we have drafted have evolved over the last several months. The standards document seems simple and obvious--but was born only after much hard thinking and sharing.

It includes criteria grouped in four categories:

What is *evidence of success*? Does it have beneficial effects to the populations for whom it was designed? Does it solve substantial educational problems?

Does it meet *quality* criteria set by the expert panels, does it have clear goals and technical excellence? Does it promote equity regarding race, gender, socio-economic background and culture?

Educational significance will include such criteria as potential to increase the educational knowledge base; build on successful strategies; usefulness to others.

Usefulness to others: Is it available and is it reasonable compared to other alternatives given the amount of resources of a school or school system?

The expert panels that are being established will apply those criteria to the programs considered for approval. The legislation described "program" in the broadest sense. Program includes curriculum materials, assessment materials and approaches, professional development programs, research findings and policies. The expert panels will use the final criteria to make a decision as to whether a submitted program is exemplary or promising.

The distinction between promising and exemplary that we are proposing is that promising programs will have evidence of success in one context or with a distinct population; whereas exemplary programs will have evidence of success in multiple contexts and with multiple populations. I especially like Major Owens' ideas about certification and trying out promising programs. This is a good idea that I hadn't heard before.

As Alba has related, the OERI Board has approved the draft that you have for publication in the Federal Register. The comment period will extend for 60 days after publication. We hope you will participate. To date we have started to develop two pilot expert panels. You will hear more about both of them later this morning. The organizers of the Mathematics-Science Panel have already decided that in the beginning, programs considered will be limited to elementary education. If you have names to recommend as possible members of either of the pilot panels, please let one of us know.

As soon as a program is reviewed and becomes designated as either a promising or exemplary program, it will become part of the National

Educational Dissemination System (NEDS). Bob Stonehill, our next speaker, will tell you about that system.

Robert Stonehill, ORAD, OERI. The National Education Dissemination System (NEDS) is not only ambitious, but environmentally sensitive (holding up page), it is summarized for you on a one page handout. I had a sensational multi-media presentation for you, but apparently we are meeting in a media-free zone and they don't have the technology to support my planned show.

Our current infrastructure for helping researchers, teachers, and others to access each others work and share success with one another, includes the educational labs, research centers, and other programs identified by Major Owens earlier. We need to develop a system for making education successes available beyond those sites. NEDS is our overarching term for strategies we are putting in place to enable educators to access particular information when they need it and in the form in which they need it. It is really a system for both high-tech, low-tech, and no-tech access. We still get lots of letters and phone calls and E-mail requests. Requests that require a human being to syphon the request and give it a response.

We are creating a self-serve service system where clients themselves can access what is needed. (Bob discussed the diagram on the handout.) NEDS is primarily a service provider. It will include an ED Web repository and will connect the best national, state, and local resources, the work of educational laboratories, national research centers, colleges and universities, government supported programs. All of these resources will be accessible to a person sitting at a service provider.

We are co-developing NEDS with many people and programs and building on past and current successes. For example, both the Department of Education and OERI have 800 numbers and both have an Internet presence. OERI has supported the development of numerous outstanding school reform programs over recent years. 1.8 million persons per month access the ED Home Page. Both the ED Home Page and ERIC have been cited among the top five best world-wide resources in education. People are constantly searching for what works best and for what research tells us about, as Eve said, fill in the blank.

For the first time we are attempting to provide open access to what we have developed over the last 30 years. We need your help in building a system that will provide access to what you and others have developed.

The education communication networks are already further along than many think. There are already many ways for educators to access information and expert assistance. Help us to develop NEDS and move these sharing systems even further along. I will close with some questions for your group to consider: What aspects of NEDS would be most useful to you? What can you offer to NEDS? Who else must be included? How can we best work in a de-centralized environment? How do we achieve the primary intents of NEDS? (--To create, adapt, identify, validate and disseminate effective and promising educational programs;--To increase the capacity of teachers to participate in research and development?)

Sue Klein, ORAD, OERI (main organizer of this session). I want to thank Major Owens, Assistant Secretary Robinson and everyone else on the panel, and everyone in the room for participating in this opportunity that AERA has given to us for bringing a group of national educational leaders together to work on these important new educational initiatives.

I want to emphasize that this new responsibility to coordinate and disseminate the best educational programs is not just for OERI programs, but for the rest of the programs administered by the Department of Education, for educational programs supported by other federal agencies, for successful programs and practices emerging from states, local school systems, educational organizations, private publishers, and interested educators no matter where or under what auspices they are developed.

The responsibility is to share what works, using a total system of topic focused expert panels, using consistent standards to identify promising and exemplary programs and practices that will truly help our education constituents, teachers, students, parents, researchers, and other users. The system's purpose is to help consumers find the best for them among the many promising and exemplary R&D based products, programs and practices. There will probably never be a one best solution, but multiple good solutions, each with its own strengths and weaknesses, and each working differentially in different contexts when adapted by users.

Ideally, we hope the system will help the education field develop consumer reports for specific topic areas where we have expert panels. Educators need ways to retrieve descriptive and evaluative reviews of needed educational products so they can make their own side-by-side, product to product comparisons using standardized information from computerized databases.

I also see the potential for the system to help OERI achieve a new kind of collaborative leadership, ala Assistant Secretary Robinson's ideas about nourishing learning communities. To find the best in particular educational areas, we will work with colleagues in a variety of topic areas to form panels of outstanding educators. We hope constituents in these fields will take ownership of the process and recommend the best from their experience.

Coming from a research agency, we also believe in the need for good research and evaluation to justify decisions about promising and exemplary and strengths and weaknesses of each. Thus, we hope that we will be able to develop easy-to use feedback systems to learn from practitioners what works well for them as well as a process to synthesize and share this feedback with potential new consumers. We hope to encourage a more public consumer-oriented approach to evaluation that is modeled on a comparative review and analysis of best bets for similar purposes—much like consumer reports. To do so, we have commissioned research papers to help us understand what other federal agencies, foundations and organizations have done to designate programs, products, practices or research findings that are most likely to help their constituents. Some of the authors (Joan Ruskus, SRI International. and Janet Carter from Bruner Foundation) are here today as are some leaders of the review efforts they are studying such as Ken Komoski from Educational Products Information Exchange (EPIE) and Ellen Myers from Impact II Teacher Network which identifies promising teacher practices.

We also hope that working with constituency groups in identifying what works in areas they care about will have significant positive by-products for OERI's other responsibilities. For example, by learning what works and where the gaps are in potential solutions for key educational challenges, the expert panels and their advisory groups can help guide OERI R&D priorities. We will also be able to learn what does not work well in specific contexts and with specific populations. For example, we would like information on indicators of success to be disaggregated by gender, race, and language proficiency so selectors can choose appropriate programs.

In conclusion, we are in initial stages of this design. We want it to be responsive to as many needs as possible. We also want it to be inclusive. Our federal role is to make it happen over many years, to find some money to do this --but to make it something that all who want will participate and all will be treated fairly and openly. We also have the ability to connect it to many other dissemination and technical assistance activities in the NEDS.

The feedback discussions we have planned for this session are one start. We invite you to join as advisors by signing up for FINDBEST listserv and letting us know if you are interested in serving on advisory groups to explore developing new expert panels in areas of your expertise.

Use the feedback sheets that you have and give them to one of us or turn them in at the OERI booth 97 in the Exhibition Hall. I would like to introduce Beatrice Birman and Senta Raizen who will give you a brief summary on our progress to date in forming the mathematics-science expert panel and they will be followed by Pat Campbell who will give a progress report on the gender equity expert panel.

Beatrice Birman, Pelavin Research Institute and the American Institutes for Research. Greetings. I work for the American Institutes for Research, a contractor that is assisting OERI both in the development and convening of the expert panels and in developing guidelines for how these expert panels will work. We are just getting started so I will review activities to date. The biggest challenge facing the mathematics-science panel is the breadth of the fields that the panels will have to make judgements about. During the first year, which is a pilot year to refine procedures, the panel will look only at elementary science and mathematics programs.

Our job will be to help develop and support the panel and to gather information about pilot panel activities to be used in preparing an "evolving guidebook" to help future panel activities. We expect that following the general parameters of the new OERI standards, each panel will design its own system for operating. Some of the main issues each will face, and we will help with, include: synthesizing and interpreting the commissioned papers, interpreting the standards for designating promising and exemplary programs, and developing strategies for obtaining what is to be judged. We anticipate two or three panel meetings the first year and that materials will be developed to engage the public in identifying key programs to be evaluated.

We also have the responsibility of developing guidelines for future expert panels by learning from the activities of the math-science panel, the gender equity panel and other related OERI activities. During this formative first year OERI will be trying to learn as much as possible from these pilot efforts, so we will work with OERI staff and key field representatives to identify issues and develop options. For example, we have identified a number of topics that need attention at the first mathematics and science expert panel meeting, including: 1) the scope of what is considered, e.g., bigger than a lesson plan but smaller than a multi-grade curriculum; 2) additional criteria to be used in making selections; 3) how to identify and determine which programs to consider; 4) how to deal with copyright issues; 5) how to respond to challenges to panel decisions; 6) how to choose other reviewers in addition to the core panel members. (One of the views of the original panel and process is that it would be something like a journal review process, i.e., materials and programs considered good enough to make the cut would be sent to expert reviewers in the field of the product. A network of field reviewers would be established, encompassing a broad range of specialties within a given educational area); 7) how to deal with products that have already been evaluated by other agencies or systems; e.g., NSF. We look forward to getting input from you regarding options and issues to be presented to the panel. We also need some trial balloons, some programs and practices to be judged at the first meeting of the panel.

Senta Raizen, National Center for Improving Science Education. Our job is not to evaluate the panel outcomes. In this initiative we are formative evaluators. Our role is to help OERI and the math-science panel by providing support and giving feedback on what we are hearing and perceiving so they (OERI) can do the job somewhat better. We commend OERI and ORAD for building in that kind of formative activity, its not usually done. We've identified seven questions that we will ask as we follow the process: **1. Who participates in the pilot mathematics and science education expert panel process?** The pilot mathematics and science education expert panel process will involve OERI staff, experts in mathematics and science education, independent contractors, staff from other agencies, potential users, and other stakeholders. How participants are selected, their roles and interactions will largely define the expert panel process. **2. How are potentially promising and exemplary products, programs and practices selected for consideration by the panel?** OERI pilot panel planning documents mention that OERI staff and independent contractors will be involved in identifying potentially promising and exemplary programs. Expert panelists will work with these nominations. The procedures used to search for and identify potentially promising and exemplary programs, (e.g., equal opportunity for identification, the comprehensiveness and thoroughness of the search, ability to range from identifying individual products to broad-scale systemic initiatives), are central to the credibility of the panel process. **3. How do participants use available resources?** It appears that participants will have a myriad of resources available during the panel process. These range from draft standards for promising and exemplary programs to an expert panel guidebook, database and commissioned papers. This evaluation question looks at participants' resources use as well as the usefulness of specific resources to assist the expert panel deliberations and decisions. Characteristics of resources not available but seen as potentially valuable are particularly important given OERI's opportunity to refine the expert panel process before large-scale implementation. **4. What are the nature and outcomes of the panel's formal and informal activities?** This evaluation question addresses the documentation and description of the actual expert panel process. Formal activities refers to panel meetings (and subpanel meetings if this option is selected). Informal activities refer to interactions between meetings. Outcomes encompasses decisions made by the panel with respect to criteria or standards for recognizing promising and exemplary programs, procedures governing panel operations, obtaining potentially promising or exemplary programs, communicating with stakeholders, choosing promising and exemplary programs, and requests and assignments to independent support contractors. The nature of activities refers to the group processes used by the panel to make decisions. **5. How do participants respond to problems encountered in the pilot process and unforseen events and opportunities?** Problems, unforeseen events, and unforeseen opportunities are a valuable source of learning. This evaluation question seeks to focus on the challenges and unanticipated consequences faced by the expert panel during the process to understand how these influence the nature and outcomes of the panel's deliberations. With each, an estimate of likelihood and potential impact can be made such that addressing the event, problem, or opportunity might be built into future expert panel procedures. **6. How do participants assess the credibility and efficacy of the pilot process?** Participant opinions as to the quality, effectiveness and usefulness of the expert panel process approximate a market test of the efficacy of this approach to the identification of promising and exemplary programs. Participants' insights into specific "defining moments' or "turning points" in the process will be particularly interesting so that these might be built into future expert panel procedures. **7. What dissemination activities are planned or implemented during the pilot process:** Disseminating the results of the expert panel process is part of OERI's pilot test of the process. Dissemination begins in the process design phase with OERI staff reaching out to associations and other groups representing stakeholders for opinions about the process design. Dissemination and dissemination planning should continue throughout the pilot test to create a readiness on the part of stakeholders to use the expert panel results.

NCISE proposes to use several methods to collect formative evaluation data in response to each evaluation question. These methods include: document review, interview, direct observation, and written questionnaire.

Sue Klein. Since Patricia Campbell is not here, I will give a brief overview on the development of the Gender Equity Expert Panel. (She was stuck in MA with 8" of snow.) Pat Campbell is the formative evaluator and Kathryn Hansen, of the Women's Educational Equity Act (WEEA) Resource Center at the Educational Development Center (EDC) in Newton, Massachusetts, is currently providing support services to the panel. The panel membership and network will be very inclusive. Evolving from two preliminary planning meetings of gender equity expert panel advisors, it was decided to focus on six issues: 1) core gender equity; 2) teacher professional development; 3) disabilities; 4) school-to-work; 5) combating violence; and 6) mathematics, science, and technology. Given its overarching mission, the panel will also work with other panels and we are building in systematic interaction with the math-science expert panel.

Sharon Robinson. I would like to thank all of the presenters for their concise and high quality presentations. They have given us an excellent overview on OERI's progress to date in developing these wonderful and important new educational initiatives. In case he has to leave before we are finished, I would like to give special thanks to Major Owens for joining us to share his marvelous historical and Congressional perspective on the new OERI! His remarks, as always were well thought out, provocative, and inspirational.

Now to the fun part for those of you in the audience. We would like to have you form four discussion groups around the topics of: teachers; developer/publishers; technical assistance providers; and researchers/evaluators. We would like your discussions to focus on what you have heard today about the new standards, NEDS system, and expert panels. Develop several key questions or issues and identify a spokesperson for the group who will share these issues with the entire group during our closing plenary discussion.

Group Presentations.

What might this mean to teachers?

Group Leaders: Ellen Meyers, Impact II Teacher Network; Virginia Richardson, University of Arizona

o Please be sure to have a healthy representation of teachers on every expert panel.

o Time is a key issue for teachers so please design a dissemination system in which teachers will be able to find and access needed products in as short a time as possible.

o Teacher interest is also a key issue so build in incentives to attract teachers to the system.

o Include the World Wide Web (WWW) in your dissemination system.

o If possible, design NEDS so that whenever possible, it relates approved programs and practices to classroom instruction.

What are the implications for R&D technical assistance providers?

Group Leaders: Bob Blum and Joan Shaughnessey, Northwest Regional Educational Laboratory; Diane Lassman, The Exchange; Ludwig Braun, International Society for Technology Education; Glenn Hultman, Linklping University, Sweden; Florence Falk-Dickler, Region 2 Women's Bureau; Cheryl Garnett, OERI

o There will be a real challenge in applying the standards to "successful" programs that are softer in results, e.g., we know they work but it is hard to prove. There should be an option to recognize some programs even without hard data.

o There are a variety of important equity issues that are not built into the standards as strongly as they should be.

o Technical assistance is a broad topic and is much more than getting a lot of good stuff to people in the field--there is a big distance from identifying products to getting them approved, recognized, accessed and used in practice.

What are the implications for developers/publishers?

Group Leaders: Charlotte Frank, McGraw Hill; Helen Farmer, University of Illinois; David Rodriguez, Educational Partnership Group; Richard Delano, Scholastic Inc.

o Timelines are important. The system should facilitate the institutionalization of good programs and practices as quickly as possible. We need change now!

o Cost is a central variable, e.g., what if a program cost millions to implement? Be sure to collect and report good cost information for each program approved.

o Find programs to evaluate that have been adopted by governments, organizations, or programs that have a large constituency--national and state adoption are best. Developers and publishers look for this.

o There is an urgency in what you are doing. We need change now. Do it soundly, but develop and implement the system as soon as possible.

(Comment from Ken Komoski: There are 1,000 software publishers who have produced over 20,000 software packages, yet when you ask teachers who use the new technologies if they can find adequate software to integrate into their curriculum they generally say no. Main reason is that many of the publishers publish the same general stuff--and most of it is not well evaluated.)

What are the implications for researchers and evaluators?

Group Leaders: Willis Hawley, University of Maryland; Edgar Epps, University of Chicago; Ken Komoski, Educational Products Information Exchange; Abbe Herzig, University of Wisconsin, Madison, formerly with Consumer Reports; Joan Ruskus, SRI; Janet Carter, Bruner Foundation; Senta Raizen, National Center for Improving Science Education; Thomas Tyden, Dalarna Research Institute, Falun, Sweden

o How will you satisfy the broad range of consumers who have an equally broad range of what they consider to be evidence of effectiveness?

o Base the system on good design principles so panel members can do an effective job of evaluating what it is given.

588

o If you wait for people to come to you (with good products), they won't. Be practical and be proactive, go out and find the best products.

o The process can help shape research priorities.

o If it is important to approve programs and products that you want widely shared, the context within which the program and/or products works is a very important consideration--more important than usual.

Response and Closing Remarks by Major Owens, Sharon Robinson, and Others

o There is a sense of urgency. So it is important that we take what steps we can as quickly as possible.

o There is a need for judgements about what works. Even with the political context we are in, we need to make those judgements.

o Regarding the timelines, we are talking about the need for implementing our pilot activities--for making our "dry run" in one year or less--events are galloping on!

o The first two expert panels, math-science and gender equity should help accelerate our dry run. They may be controversial, but today, everything is controversial.

o How do we deal with controversy, we cannot escape such issues as privatization, charter schools, and vouchers.

o What if we cannot find any exemplary programs?

o The market place is racing ahead. In New York, everyday, some new software is introduced somewhere. The production of new products and approaches is way ahead of the evaluation of marketplace programs.

o What about private sector programs that have not gone through a tough evaluation process, programs like "Hooked on Phonics." There will be pressure by the private sector to approve and advocate such programs, many are sold through commercial advertising.

o We need to identify good products; have expert panels evaluate, approve and codify them; and then we need to disseminate them as widely as possible.

o We need to do all of the things we have been talking about, but in as thorough and scientific process as possible!

Meeting Handouts:

Session Agenda: OERI Leadership Efforts in Designing Approaches to Identify and Share Promising and Exemplary Products, Programs and Practices

Draft 4/03/96 Department of Education "Standards for Conduct and Evaluation of Activities Carried Out by the Office of Educational Research and Improvement (OERI)--Designation of Exemplary and Promising Programs"

The National Education Dissemination System (NEDS) Under Construction

Highlights of the Proposed Consumer-oriented Evaluation Component of the National Education Dissemination System, NEDS

Commissioned Papers and Formative Evaluations to help the Office of Educational Research and Improvement (OERI) in the U.S. Department of Education Design: Standards for Designating Promising and Exemplary Programs and A System of Expert Panels to use these Standards.

An Invitation to Participate in OERI's "Findbest" Listserv

V

November 19, 1996

RESPONSES TO: Options in Discussing Distinctions Between Promising and Exemplary Products, Programs, Practices, Policies and Research Findings for a System of Expert Panels

Summary of The Response Process

We posted four options on Nov. 4, 1996 on the following listservs: "findbest" for colleagues interested in advising OERI on the emerging system of expert panels, a list of the Mathematics & Science Expert Panel members, and "gndrpan" for Gender Equity Expert Panel members and advisors. We received responses to all lists and some private responses by e-mail and telephone to Drs. Sharon Bobbitt and Sue Klein in OERI. Most of the responses were directly related to the distinction between promising and exemplary. A few were only tangentially related as they were about specific resources. There were 25 on target responses by Nov. 15. The responses brought out many issues that need to be attended to in explaining and managing the work of the expert panels and provided an opportunity for respondents to share their ideas of what the expert panels should be doing. We also had some discussions on Nov. 16, 1996 with the Mathematics and Science Expert Panel on this topic during their meeting in Washington, DC.

It is likely that we could obtain more insights, guidance and even agreement if we had more discussions with respondents. We did this in one case where a respondent said he answered too hastily and would change his advice, but he wanted to be part of the conversation so he wrote a quick e-mail message. Klein and Bobbitt didn't intervene with follow-ups in listserv conversations to avoid biasing the exchange--but in the future some follow-ups might be helpful to keep the conversations focused on the key topic. It would also be helpful to allow more days for the exchange. We were pleased that many people took the time to respond and hope that many other "readers" were informed by the query and discussion. We are sharing this summary to continue the exchange. Since we combined many responses, we hope that you won't mind if we incorporated your ideas and sometimes your suggested wording without attribution. The OERI Board will be meeting on the recommendations in early December.

If you have comments on these summary recommendations or would like a copy of the initial posting on the options, please contact Sue Klein (sue_klein@ed.gov) Tel. 202-219-2038, or Sharon Bobbitt (sharon_bobbitt@ed.gov) Tel. 202-219-2126, Fax 202-219-1407, OERI, ORAD, Room 508, 555 New Jersey Ave., NW., Washington, DC 20208. Please provide your e-mail, fax or postal mail address. We can also send you a copy of the draft regulations or the pertinent sections of the OERI authorizing legislation.

This summary document includes:

I. Suggested Substitute Wording for the Section of the Draft June 3, 1996 Regulations on Differences between exemplary and promising programs.

II. Discussion of the Revised Distinctions Between Promising and Exemplary Programs for a System of Expert Panels

III. Information on How the Feedback Helped us Decide on the Winning Options

I. Suggested Substitution for Federal Register/ Vol. 61, No. 107, June 3, 1996, Proposed Rules, Page 27993 (Draft November 18, 1996).

§701.21 **What is the difference between an exemplary and promising program?**

(a) A panel may recommend to the Secretary that a program be designated as promising if the panel determines that the program is acceptable on each of the four categories of standards in §701.22 (Evidence of Success / Effectiveness, Quality of the program, Educational significance and Usefulness to others). Under evidence of effectiveness, a promising program must be judged as having a pattern of evidence that the program has worked or achieved an acceptably significant claim related to a desirable/meaningful performance outcome in at least one site and that the program has logical or other evidence of adaptability and transportability to other sites. Each panel will establish their own criteria and decision framework on acceptability for designation as promising on the three standards categories of quality, significance and usefulness to others. Panels may approve a program as promising if some, but not all criteria within any of these standards categories are judged acceptable.

(b) A panel may recommend to the Secretary that a program be designated as exemplary if the panel determines that the program is excellent on each of the four categories of standards in §701.22 (Evidence of Success / Effectiveness, Quality of the program, Educational significance and Usefulness to others). Under evidence of effectiveness, an exemplary program must be judged as having convincing evidence that important claims related to desirable/ meaningful performance outcomes can be reasonably attributed to the program, and that this pattern of success was evident in multiple site replications within the past few years. Each panel will establish their own criteria and decision framework for designation as exemplary on the other standards categories of quality, significance and usefulness to others. An exemplary designation means that the panel judges the program excellent on almost all criteria in all four standards categories and agrees that any weaknesses must be minimal and easily corrected.

II. Discussion of the Revised Distinctions Between Promising and Exemplary Programs for a System of Expert Panels

Background:

In the June 1996 Draft of the Regulations, the current standards are divided into four categories of effectiveness, quality, significance and usefulness to others with specific criteria under each. The only distinction on promising and exemplary was made for one aspect of evidence of effectiveness--replicability in single sites for promising or multiple sites for exemplary. The revised version recognizes that distinctions on promising and exemplary should be made within each of the four standards categories and that additional aspects of effectiveness should be used in these important distinctions. The revised distinctions are also based on a clearer understanding of the purposes of the entire system and how federal decision makers and others will use decisions about promising and exemplary differently in their resource allocations.

The New Version:

Since evidence of effectiveness is particularly critical to the purposes of this system of expert panels, expert panels in all topic areas will use the same decision framework to distinguish between promising and exemplary on effectiveness. However, since there are many topic area differences in criteria related to the other three standards categories of: significance, quality, and usefulness to others, each expert panel will establish its own criteria and decision framework for determining adequacy in each of the three categories for **promising** and on excellence for an **exemplary** designation.

An **exemplary** designation means that the panel judges the program excellent on almost all criteria in all four standards categories and agrees that any weaknesses must be minimal and easily corrected. Some criteria that are required for an exemplary recommendation may not be considered essential for a **promising** designation but the panels must still judge a promising program adequate on all four standards categories. (Using the June 1996 Draft OERI Regulations definition, program includes products, programs, practices, policies and research findings. However, the primary focus of this paper is on products and programs. As Michael Scriven and others have pointed out, other considerations will be needed in distinguishing between promising and exemplary research findings, policies and practices.)

Decision Framework on Evidence of Effectiveness:
For evidence of effectiveness, a designation of **promising** means that:

- there is some pattern of evidence that the program has worked or achieved an acceptably significant claim related to a desirable/meaningful performance outcome in at least one site and

- the program has logical or other evidence of adaptability and transportability to other sites.

At the minimum, the evidence of effectiveness should make a logical case that the claim of success can be attributed to the program and that it is not wholly idiosyncratic to the site unique circumstances. Thus, this evidence may come from experiences in more than one site. Qualitative information and logical analyses can be used to support the claim that it contributes substantially to a desirable outcome. (Such an outcome should be more than "the participants liked the program".)

For the effectiveness category, a designation of **exemplary** means:

- that there is convincing evidence that important claims related to desirable/ meaningful performance outcomes can be attributed to the program,

- that this pattern of success was evident in multiple site replications within the past few years and (Replication means that the common elements of a program are used in another site.)

- that the panel agrees that the defensible important claims for this program are as good or better than the defensible claims of other exemplary programs with similar purposes. (Actual comparative data collections and evaluations using the same appropriate instruments, third party evaluations, etc. are not essential for this judgement, but would be welcome.) [This item has been omitted from the revision of the regulation, but it is assumed that the panel would make this comparison anyway.]

Qualitative evidence may be used to build a case for evidence of effectiveness to support these claims. Credibility of positive claims can be strengthened by providing evidence and explanations of ineffective as well as effective replications.

Justification and Discussion of Issues:

How the New Distinctions Support System Purposes

In clarifying distinctions between promising and exemplary judgments, it is necessary to provide more specific purposes for both the overall system of expert panels and for treating promising and exemplary programs differently.

- Overall System Purposes Are:

  1) to help educators learn about effective and replicable programs so they can make informed selection decisions. It is not a recognition system (for excellence, hard work, etc.). Consumers will need to have adequate information about strengths and weaknesses of what is designated as promising and exemplary so they can make wise choices to fit their own needs and

  2) to help federal sponsors and others make more strategic and cost effective decisions about their follow-up investments in the evaluation, continued development, and dissemination of R&D based programs and to identify gaps where they should support new R&D.

- Purposes for Distinguishing Between Promising and Exemplary:

  A key purpose for designating a program "**promising**" is that it clearly merits additional investment for:
  1) *improving it* and
  2) *evaluating it* so that it could justify a future exemplary designation. Using this purpose, a promising designation would encourage only cautious adaptation or adoption—while undergoing careful evaluation. Users should be encouraged to select exemplary rather than promising programs unless no exemplary programs meet their need or they have good reasons to participate in the field testing of promising programs.

  The purpose of having an **exemplary** designation is to increase the use of what works best. Thus, users should be encouraged to choose from among the exemplary programs. In addition to multiple types of dissemination support ranging from helping with world-wide web information and publicity via tv and print media, ED and others may provide incentives and encouragement for choosing exemplary programs and should do what they can to make sure that these exemplary programs are accessible to all who could benefit from them. Since educational improvement is dynamic, routine performance information should be collected on exemplary programs and fed into a national feedback system to make sure that quality assurances (and comparative advantages) associated with exemplary designations are maintained. As needed to justify federal funding or marketing decisions, comparative evaluations of exemplary programs should be performed. The results of these comparative evaluations should be shared with others to learn if the exemplary programs have maintained their high rankings on key criteria or if they perform well on important new criteria.

Incentives:

If there are real incentives for additional support, then being judged promising would be an honor with real rewards for working toward improved programs. The formative evaluation report (for the gender equity expert panel (Campbell & Bachmann, Nov. 1996) stressed the need for supporting evaluations of programs in this area. The focus on continued evaluation of promising programs is specified in the legislation. Michael Scriven said that people should use promising programs "if no exemplary program meets their needs, although very cautiously. It is not correct to say that one should use them if one "likes" them more. The obligation is to serve recipients, not one's own likings."

The OERI Board, the pilot Expert Panels, Michael Scriven and others are aware of and want to encourage continual evaluations and improvements of education programs. Thus, they don't want to have the designations promising and exemplary be seen as good forever. One suggestion is to have a time limit, but this is difficult since it is hard to know how fast an area will change. Instead, the Board recommended that all approvals be clearly dated and the users informed that older approvals maybe superseded. The above call for comparative evaluations of exemplary programs can be one way to obtain fresh consumer information.

Importance of Focusing on Evidence of Effectiveness

These revised distinctions between promising and exemplary programs continue to emphasize the importance of examining evidence of effectiveness. This is important because we know that the most commonly omitted criteria in most education review efforts and consumer decision making is any judgment based on information on outcomes. Without this evidence, it will be impossible to know if, and with whom, a program works. By making some evidence a requirement for promising programs ED will be providing a powerful incentive for the collection of this evidence. It will also show that it is not in the business of reviewing the many documents, videos, etc that are used to share information, but not directly intended to change educator or student outcomes. This revision also allows some flexibility among the panels in choosing what is acceptable on the other three standards categories where there is high likelihood of differences within expert panel topic areas.

Issues to consider when using these distinctions to designate programs as promising or exemplary:

- A few respondents to the discussion of options on ways to distinguish between promising and exemplary stressed the importance of relying on evidence of effectiveness with attribution and warned against overweighting criteria related to quality and usefulness to others because these criteria could be subjective and tied to consensus fads of the moment about what is a good way to do things. For example, Lois-ellin Datta said "ED does not want to canonize current beliefs, however but rather to open the door to iconoclastic ideas for promoting good outcomes for learners."

- Some of the respondents from the Gender Equity Expert Panel listserv asked about and even suggested types of gender equity outcomes and underlying assumptions to be valued. The Nov. 15 and Nov. 16 discussions of the Mathematics and Science Expert Panel also started to address their selection of values such as if and how they should use national mathematics and science standards and benchmarks in their criteria.

- We need to use the many acceptable ways of making the case that claims related to attributions of effectiveness and replicability can be substantiated without total third party evaluations and control groups with rigorous evaluation on fair measures. Lois-ellin Datta says there are many ways to get good evidence "by (a) ruling out rival explanations and/or

(b) ruling in what influences what" — "However, every method has its standards of excellence and the panels would know what these are and demand these be met."

- We need to be able to talk about replicability and transportability in realistic terms that acknowledge differences in users and the important role of careful adaptation. Based on feedback, we have decided to omit the predictive term "generalizability" and instead urge submitters to make claims about patterns of performance outcomes achieved with specific populations. For example, Robert Stake said that "the state of determining what is generalizable under what conditions has not advanced sufficiently to justify using the term for purposes proposed here—." In the profiles and consumer reports describing the promising and exemplary programs, information would be provided on the types of sites and populations that achieved the claimed results, so that users could decide if it was also likely to work for them.

III. Information on How the Feedback Helped Us Decide on the Winning Options

The feedback we received helped us clarify the purposes so that the distinctions between promising and exemplary would contribute to these purposes.

Feedback on Purposes:
We learned that there was substantial confusion on the distinctions between promising and exemplary because respondents had different understandings of the purposes of the overall system and why panels should differentiate between promising and exemplary. Thus, we needed to be clearer about the overall purposes for the system of expert panels. Respondents were not satisfied with the statement in draft regulations or in the original options memo. *"Two major purposes are: 1) to share information on what works with consumers so they can make informed choices based on sound evaluation information from knowledgeable reviewers and 2) to help federal sponsors and others understand where the gaps are to guide funding of future R&D."* The more explicit revised purposes are described in Section II.

Advice on the Winning Options:
Most respondents preferred Option 3 which allows for distinctions between promising and exemplary within each of the four standards categories. They supported the emphasis on evidence of performance effectiveness and having all panels use the same distinctions for promising and exemplary evidence of effectiveness. A key distinction, preserved from Option 1 is that it must have evidence of effectiveness in at least one site for designation as promising and in more than one site for designation as exemplary. However, there was agreement that each expert panel should set their own minimums for designations of promising and exemplary in the standards categories for quality, significance, and usefulness to others. No one advocated using Option 4 where a promising designation could be given despite unacceptability on any one of the standards categories. Several respondents appreciated the flexibility in option 2 that the panel members would decide on cutoffs within the standards categories according to a continuum of excellence. This has been included in the revision of Option # 3 by giving individual panels this responsibility on all standards categories except effectiveness, but It is more specific than Option 2 because it requires the panels to be clear about minimums for promising and for exemplary in each standards category.

W

Expert Panels for Promising and Exemplary Innovations: A Fine Idea From the Feds [*]

by
Peter Seidman
Dissemination Program Director
National Center for Research in Vocational Education
University of California, Berkeley

[* This article is based on two electronic messages submitted to NCRVE's DISSMN8 list, posted by Sue Klein of the US Department of Education, Office of Educational Research and Improvement. The ideas, points of view, and opinions expressed in this *CenterWork* article are those of its author. They do not necessarily represent Dr. Klein's or the official U.S. Department of Education position or policy including, but not limited to those of its Office of Educational Research and Improvement and the USDED Office of Vocational and Adult Education. All information in this article concerning OERI's initiative is accurate as of this writing, but is not the final word given the initiative is still very much under development and review.]

While national systems whose task is to share information about education are alive and well (for example, ERIC, the Internet, the World Wide Web), this past year has witnessed the demise of several national educational dissemination programs. These programs were part of a national infrastructure that helped state and local educators to identify and replicate successful materials and programs in their areas. Among the "casualties" were the National Diffusion Network (NDN), the National Network for Curriculum Coordination (NNCC), and the Regional Centers on Drug-Free Schools and Communities. While pieces of this infrastructure still remain, the programs no longer exist as national systems for disseminating data, knowledge, and information. And their absence is sorely missed.

For example, the most inquired about topic when folks call the National Center's Dissemination Program is for exemplary curricula and programs. And the question, "What works?" is arguably the most frequently asked question in education. When the NNCC as well as the NDN existed, our information brokers had national systems to which to refer the callers.

Another example comes from a colleague who worked at one of the defunded Regional Centers for Drug-Free Schools and Communities. Each day now she is inundated with calls to her home from frustrated people who have no national system from which to gain needed information. Sadly, she lacks the time and resources to assist these callers.

---

[*] Reprinted from *CenterWork* 8 (1), 5-6.

Looking into the near future, state and federal governments may find that they need the very infra-structure that has recently been dismantled through zero funding. Ironically, they may end up having to rebuild these same systems.

Filling the Infrastructure Gap

Fortunately, the U.S. Department of Education's Office of Educational Research and Improvement (OERI) is now working to fill a part of this vacuum. OERI was mandated by its 1994 reauthorization to develop methods to identify and share both promising and exemplary materials and programs--or what we shall refer to in this article as "educational innovations." OERI seeks an approach to dissemination and replication that is realistic, acknowledging differences in users and the important role of careful adaptation.

In order to carry out its mandate, OERI is establishing a system of expert panels to carry out this mandate. Their job is (1) to help educators learn about effective and replicable innovations so they can make informed choices; (2) to help federal sponsors and other funding organizations make more informed and cost-effective decisions when considering refunding programs for evaluation, continued development or dissemination purposes; and (3) to help such sponsors identify gaps in research and development where new support would be useful.

OERI, which was responsible for maintaining the now extinct NDN, favors the expert panels approach. In the past, NDN utilized a single, independent panel which passed judgment on which programs should receive exemplary status. OERI envisions expanding this idea into a system of expert panels on various topics, established by experts in those areas. OERI would assist panels in the coordination and facilitation of their activities, it would encourage the panels to design many of their own activities to address the needs and opportunities in their topic area.

One of the main charges of the expert panels would be to distinguish between those programs considered "promising" and those considered "exemplary." To do so, evaluators would consider four criteria or standards for educational excellence:

1. Evidence of success/effectiveness
2. Quality of the innovation
3. Educational significance
4. Usefulness to others

"Promising" Versus "Exemplary" Purposes

A key purpose for designating a program promising is that it clearly merits additional investment for (1) improving it, and (2) evaluating it so that it could justify a future exemplary designation. Using this purpose, a promising designation would encourage only cautious adaptation or adoption, during which time it would undergo careful evaluation. Users would be encouraged to select exemplary rather than promising innovations unless no exemplary innovations meet their need, or unless they wish to participate in the field testing of promising innovations.

The key purpose for designating a program as exemplary is to increase the use of what works best. Users would be more likely to adopt as their model a program that has received the exemplary designation. OERI plans to assist in publicizing the existence of exemplary programs through the World Wide Web, and TV and print media.

"Promising" Versus "Exemplary" Criteria

In recent electronically transmitted requests for comments, OERI first laid out several options by which this system of expert panels could discuss distinctions between promising and exemplary innovations. These requests were sufficiently provocative and of sufficient import to summarize them here.

As you read the following ideas, imagine that you have an innovative program or product which you believe is of such high quality that it would be of genuine use to others. Imagine further that a federal program exists to identify and disseminate promising and exemplary innovations such as yours.

Finally, imagine that your innovation is to be scrutinized by a federal expert panel. Perhaps it has to do with school-to-work, technology education, workplace mentoring, or business/community/school partnerships. Which of the following four options seem most viable to you? Why?

A description of the four options follows, as does the choice OERI has made (Option 3) that it will recommend for the Department of Education regulations on these standards. Nevertheless, as you reflect on the substance of the article--and if you are sufficiently interested--we have included contact information at this article's end so that you, too, can offer your views to the Department.

The Four Options

In its first electronic posting, OERI laid out four options that the expert panels might use when distinguishing between promising and exemplary innovations.

Option One: **Distinction Based Only on Effective Replication.**

A program may be designated as promising if the panel determines that the program has met the educational effectiveness standard with respect to one context or one population. A program may be designated as exemplary if the panel determines that the program has met the educational effectiveness standard with respect to multiple contexts or multiple populations.

Option Two: **No Clear Minimum Expectations for Either Promising or Exemplary on any of the Four Standards of Educational Excellence.**

A continuum of excellence would be established by each panel for each of the four standards. No system-wide minimum expectations would be established for any of the four standards for either promising or exemplary innovations. Those innovations rated lower on the four standards categories would be called promising. Those rated substantially higher would be called exemplary.

Option Three: **Focus on Evidence of Effectiveness, but Establish Minimums for Promising and Exemplary on the Other Three Standards of Educational Excellence.**

An innovation must attain minimum expectations on all four standards to be called promising. Each panel will decide on its own minimums for all categories except that of evidence of effectiveness, for which systemwide minimums will be established. To be exemplary requires an "excellence" rating on all four standards categories as determined by each expert panel.

Option Four: **Number of Categories Judged Excellent: Exemplary Requires Four; Promising Requires Only Three.**

Exemplary requires an "excellence" rating on all four standards categories as determined by each expert panel. Promising requires an "excellence" rating on at least three standards categories. Thus, any innovation may be approved as promising even if no evidence of effectiveness exists (that is, if the innovation is rated excellent on the other three standards categories).

Why OERI Chose Option 3

Most respondents to OERI's electronic posting preferred a revised Option 3, which allows for distinctions between promising and exemplary within each of the four standards categories. They supported the emphasis on evidence of performance effectiveness, and also liked having all panels use the same distinctions for evidence of effectiveness. A key distinction, preserved from Option 1, is that the educational innovation must demonstrate evidence of effectiveness in at least one site in order to be designated as promising, and in more than one site to be designated as exemplary.

Most respondents agreed, however, that each expert panel should set its own minimums for designations of promising and exemplary in the standards categories for quality, significance, and usefulness to others.

Additionally, the flexibility in Option 2, which empowers panel members to decide on cutoffs within the standards categories according to a continuum of excellence has been integrated into Option 3 by "popular demand" of the respondents. Individual panels will be given this responsibility on all standards categories except effectiveness. Additionally, the revised Option 3 is more specific than Option 2, because it requires the panels to be clear about minimums for promising and for exemplary in each standards category. (For the complete text see OERI's proposal at the end of this article.)

Importance of Evidence of Effectiveness

These revised distinctions between promising and exemplary innovations continue to emphasize the importance of examining evidence of effectiveness. Such emphasis is important because effectiveness is the most commonly omitted criterion in most educational review efforts and consumer decision making. Without such evidence, it is impossible to know if, and with whom, a program works. By making such evidence a requirement, the Department of Education will be providing a powerful incentive for its collection.

Since evidence of effectiveness is particularly critical to the purposes of this system of expert panels, expert panels in all topic areas will use identical effectiveness criteria to distinguish between promising and exemplary programs. However, since there are many topic area differences related to the other three standards categories, each expert panel will establish its own criteria for determining adequacy in the case of promising designations and excellence in the case of exemplary ones.

Exciting Possibilities for Voc Ed Expert Panels

OERI staff has made it quite clear that they welcome vocational education and school-to-work educators to design and operate expert panels to identify and share promising and exemplary innovations. Such panels can branch off into subpanels around domains within the purview of the full panel's charge. For example, within a school-to-work expert panel one could have subpanels encouraging submissions of innovations in school-based learning, work-based learning, and connecting activities. There could be subpanels on tech prep, curricular integration, business/industry/education partnerships. Without question, vocational education, school-to-work, and workforce development all have promising and exemplary innovations worthy of government support.

The question is from where would such public support come? How about legislating it? The School-to-Work Opportunities Act (STWOA) mandates or allows a host of school-based and work-based activities that could evolve into promising and exemplary innovations. Could STWOA monies be used to support expert panels to identify promising and exemplary innovations? How about Perkins funds? Could resources from either of these federal laws be used to establish a pilot project in which an expert panel on a single topic, such as workplace mentoring, would be established? Could the panel's creation and operation then be studied for use in writing future legislative language? And what about future workforce development legislation? Could resources be allocated for expert panels not only to identify promising and exemplary innovation, but also to support continued development of promising ideas and sharing of exemplary ones?

OERI has a solid vision, and one we believe is worth committing oneself to: The creation of federally supported expert panels "owned" by the constituency groups expert in the panel's selected topic, whose purpose is to identify promising and exemplary innovations within their area of expertise. Once identified, resources would exist so that exemplary products and programs could be shared. Equally important, promising innovations could be evaluated and revised into exemplary status--an unusual approach for a federal program, but one with great possibilities.

Thus, OERI is offering us a vision, one which it asks us to accept. It believes that identifying and sharing promising and exemplary innovations is important to the improvement of American education. Simultaneously, the OERI vision allows constituency groups--such as vocational educators--to own the identification and dissemination systems, and to customize these systems to the idiosyncrasies of the topic and those who will use them.

Having your cake and eating it too? Perhaps, and it sounds worth the gamble.

ADDENDUM

TEXT OF OERI PROPOSED LANGUAGE

§701.21 What is the difference between an exemplary and a promising program?

a. A panel may recommend to the Secretary that a program be designated as promising if the panel determines that the program is at least acceptable on each of the four categories of standards in □701.22 (Evidence of Success / Effectiveness, Quality of the program, Educational significance and Usefulness to others). To be judged acceptable in the category of evidence of effectiveness all of the following criteria must be met. A promising program must:

- have defensible overall evidence supporting claims of worthwhile performance results (without substantial harmful results) at one or more sites (without failing at a large number of other sites)
- have logical or other evidence of adaptability or transportability to other sites.

The combination of this positive evidence of effectiveness and potential replicability creates a significant probability that the program will eventually be able to provide evidence to support claims of exemplary meritorious results as defined in the following section (b).

b. A panel may recommend to the Secretary that a program be designated as exemplary if the panel determines that the program is excellent on each of the four categories of standards in □701.22 (Evidence of Success / Effectiveness, Quality of the program, Educational significance and Usefulness to others).

To be judged excellent under the category of evidence of success/ effectiveness all of the following criteria must be met. There must be convincing evidence that very important claims of positive results (or performance outcomes) can be:

- reasonably attributed to the program, and
- that evidence to support these claims of worthwhile results was sustained in multiple site replications within the past few years (without failing at a large number of other sites or being accompanied by harmful results).

§701.22 What criteria are used to evaluate programs for exemplary or promising designation?

In determining whether an educational program ("program" includes educational polices, research findings, practices and products) should be recommended as exemplary, promising, or neither, each expert panel shall consider the following four criteria categories: evidence of effectiveness/ success, quality, educational significance and usefulness to others. Each panel must address all criteria specified in (a) below for the category of evidence of effectiveness/ success.

On the other three standards categories of quality, significance and usefulness to others, each panel may add to the core criteria and establish its own decision framework for designation as promising and exemplary. It is expected that the panels will require an acceptable judgment on almost all criteria within each of these three categories for a designation as promising and a judgment of excellence on almost all criteria within each of the three categories for a designation as exemplary. For a program to be designated exemplary any weaknesses must be minimal and easily corrected.

a. Evidence of effectiveness/ success.
  1. To be judged acceptable in the category of evidence of effectiveness all of the following criteria must be met. A promising program must:
    + have defensible overall evidence supporting claims of worthwhile performance results (without substantial harmful results) at one or more sites (without failing at a large number of other sites)
    + have logical or other evidence of adaptability or transportability to other sites.

  The combination of this positive evidence of effectiveness and potential replicability creates a significant probability that the program will eventually be able to provide evidence to support claims of exemplary meritorious results as defined in the following section.

  2. To be judged excellent under the category of evidence of success/effectiveness all of the following criteria must be met. Thus an exemplary program must have convincing evidence that very important claims of positive results (or performance outcomes) can be:
    + reasonably attributed to the program, and
    + that evidence to support these claims of worthwhile results was sustained in multiple site replications within the past few years (without failing at a large number of other sites or being accompanied by harmful results).

b. Quality. The panels will make their judgments about quality by reviewing the program materials and determining the extent to which the program:
  1. is congruent with sound research and practice
  2. incorporates accurate and up-to-date information/content
  3. promotes equity and is free of bias based on race, gender, age,

        culture, ethnic origin, disability, or limited English proficiency
    4. is appropriate, engaging, and motivating for the intended audiences
    5. contains materials that conform to accepted standards of technical product quality.
- c. Educational Significance. The panels will use their expertise in the area to determine the extent to which the program:
    1. addresses an important education issue, challenge or problem
    2. has advantages over other programs with similar purposes
- d. Usefulness to Others. The panels will make these judgments by using their knowledge of what is valued by educator and student users in determining the extent to which the program:
    1. is reasonable in terms of costs to potential users in relation to expected benefits.
    2. is or can be made easily available to potential users
    3. can be readily adopted or easily adapted in new locations
    4. can be used in conjunction with other programs if appropriate

Hello NCRVE?
Could you give me something
to use in my classroom
to help my kids learn better?

Yes, we can help you with that.
Would you like promising somethings
or exemplary somethings,
because promising somethings are okay
and could be improved in the future
until they become exemplary somethings,
but exemplary somethings
are already exemplary,
so they are what we'd recommend.

I'll take the exemplary somethings then.

You'll be happy with this choice.
Our exemplary somethings work,
a feature we often overlooked in the past,
but now we make sure our somethings
not only work, but really work,
and they didn't only work that one time,
but could work for you.

When can I expect my somethings?

Just as soon as they pass the panel of experts
who decides on whether
it is a promising something
or an exemplary something,
which will happen as soon as they decide
how they will decide, and after they decide
on their definition of effectiveness, and how
they can decide if something could *really* be
effective for others, and what the logic of
that decision is, and if, in the end, they will
need to submit the somethings to formal
third-party evaluations and control groups.

Could you send them to me priority mail?

Sure!

X

Design Competitions: A Proposal for a New Federal Role in Educational Research and Development

ROBERT E. SLAVIN

WANTED:
Instructional programs capable of significantly raising student achievement on tests linked to national standards. Federal grants available to qualified developers. Program outcomes will be compared to current methods by independent evaluators. Proposals welcome from all sources, all philosophical and theoretical bases. Put your best to the test!

For decades, policymakers have complained that the federal education research and development enterprise has had too little impact on the practice of education (see, for example, Vinovskis, 1993). With a few notable exceptions, this perception is, I believe, largely correct. Federally funded educational R&D has done a good job of producing information to inform educational practice, but has created few well-validated programs or practices that have entered widespread use. The limited direct influence of federal educational R&D, compared to that of, say, research in medicine, physics, and chemistry can certainly be ascribed in part to the far more limited federal investment in educational R&D, coupled with federal policies opposing investment in curriculum development dating back to the Nixon administration and a conservative backlash against such values-laden curricula as "Man: A Course of Study" in the 1970s.

However, in recent years the situation has changed in several important ways. The national standards movement has helped define high expectations for all children and in many ways has raised the stakes for education reformers by insisting on authentic student performance (variously measured) as the goal of reform. Adoption in many states of accountability systems based on performance assessments, a movement promoted by recent changes in Title I legislation, gives educators widely accepted performance goals worth aiming for.

At the same time, there are very important developments taking place in school-by-school reform almost entirely outside of the federal R&D structure. I am referring to the creation of ambitious, comprehensive models of school reform supported by national networks of staff developers and participating schools, such as James Comer's (1988) School Development Program, Henry Levin's (1987) Accelerated Schools, Ted Sizer's (1984) Coalition of Essential Schools, and our own Success for All and Roots and Wings programs (Slavin, Madden, Dolan, & Wasik, 1994, 1996; Slavin et al., 1994). Each of these networks includes hundreds of schools, regional training programs, and efforts to build collaboration and concern for instructional quality within and among a vast and geographically dispersed set of schools. Reading Recovery (Pinnell, DeFord, & Lyons, 1988), which is not a schoolwide change design but has had profound effects on educational practice, maintains a network involving thousands of schools. More recently, the New American Schools Development Corporation (NASDC), now called New American Schools (NAS), funded the development and dissemination of seven schoolwide reform designs. Three of these were built on existing national reform groups: the Atlas design was primarily based on Comer's and Sizer's programs, the National Alliance for Restructuring Education was based on the New Standards Project, and our own Roots and Wings model (Slavin, Madden, Dolan, & Wasik, 1994) was based on our Success for All program. However, the development and dissemination agendas of these groups have been substantially advanced by NAS support, and seven promising reform models have been added to the nation's supply of school-reform initiatives. These models are already in use in hundreds of schools. To these might be added the Carnegie Corporation's Middle Grade School State Policy Initiative—which is reforming middle schools in 15 states—the College Boards's Equity 2000 project, curriculum-specific professional development networks such as the National Writing Project and the National Council of Teachers of Mathematics, and many smaller but still impressive school-reform networks.

There is one striking commonality to all of these school-reform networks. With the sole exception of our own Success for All program, none of these networks has benefited in any important way from the federal R&D structure. In our own case, Success for All did begin in what was then our Center for Research on the Education of Disadvantaged Students at Johns Hopkins University, funded by OERI, and most of the development and evaluation of Success for All was funded by OERI through our center grants. We have also benefited by collaboration with the Southwest Regional Laboratory (SWRL), now part of WestEd, which has created a regional training program for Success for All in California, Arizona, Nevada, and Utah. The U.S. Department of Education's National Diffusion Network (NDN) has provided important support to Reading Recovery and our cooperative learning programs, among many others. However, no other major reform network has been developed or disseminated by any OERI-funded center or lab, and even in our case, our development and research has also depended in large part on grants from NAS and from the Carnegie and Pew Foundations and local Baltimore foundations. While one OERI-funded laboratory has supported our efforts, the other nine have shown little interest in doing so, and to my knowledge, no lab has either developed or disseminated any reform model on the scale of these national reform networks. Instead, national reform networks have depended on foundation and corporate grants for their development, evaluation, and dissemination.

I am not claiming that each of these reform models is ideal or that the system that supported and nurtured them through the development and dissemination process is optimal. In particular, many of the reform models have gone to national dissemination before building even a rudimentary research base showing that the program produces the outcomes it claims. Maintaining implementation quality on a large scale is a problem faced by all large professional development networks (see Stringfield & Herman, 1995). However, the success of these national programs does demonstrate several important things. First, there is an enormous hunger for comprehensive reform in schools. Collectively, these and other reform networks serve thousands of schools. Most (including ours) depend for their dissemination on charges to schools, yet many schools are willing to pay these costs, especially schools with high Title I allocations. Most require substantial buy-in on the part of school staffs; our own programs require a vote of 80% of school staff by secret ballot, and most of the other NAS designs use a similar procedure. Yet this degree of buy-in is not difficult to obtain. While funding for ambitious reforms is limited, especially in non-Title I schools, a greater limitation on the diffusion of most national reform models is a shortage of training capacity, particularly staff to do professional development and follow-up (and unwillingness or inability of many reformers to manage massive training staffs), rather than a shortage of demand. Second, the national reform programs have learned how to build efficient, durable, largely self-sustaining networks of training, follow-up, and mutual support among participating schools. Many of them have scores of examples of schools that have actively implemented their programs for as long as 5 to 10 years.

Many policymakers and reformers have advocated systemic reforms, such as adoption of standards, accountability systems, and governance changes, expressing doubt that school-by-school change can take place on a large scale (see Smith & O'Day, 1991). While systemic reform advocates note the importance of professional development in any comprehensive reform strategy (Goertz, Floden, & O'Day, 1996), they also note the poor quality of most professional development and limited transfer to teachers' daily behaviors (Corcoran & Goertz, 1995). The national school reform models provide the best available evidence that school-by-school change can take place.

If the national reform networks are doing so well without much federal assistance, why should the federal government get involved in supporting such efforts? There are several reasons. First, although these networks involve thousands of schools, there are tens of thousands of schools in need of more effective practices untouched by any systemic reform program. Even among schools participating in these networks, there are many—sometimes an openly acknowledged majority—that are very far from achieving the program's ambitious goals, and in some (as in the case of Reading Recovery and other single-curriculum reforms), the program's goals leave a great deal of the school untouched. Second, as noted earlier, most of these national models have limited evidence of effectiveness, and schools have little basis on which to choose one over another (or over doing nothing). This means that faddism or salesmanship can influence the program-adoption process more than evidence of effectiveness (see Slavin, 1989). Third, federal involvement is necessary to consolidate and institutionalize the process of developing, evaluating, and disseminating effective programs. Pioneers may blaze the first trails, but the government builds the highways. The national reform models have succeeded at the margins of federal and state policies, sometimes *despite* these policies, yet it is foolish and inefficient for these efforts to remain on a parallel track to federally supported R&D efforts and federal education funding programs such as Title I, professional development, and technical assistance programs.

Design Competitions

I would propose a radically different approach to educational research and development, not to replace existing centers and field-initiated studies sponsored primarily by OERI, but to supplement these efforts. The key idea in this approach is an emphasis on design competitions. In design competitions, government agencies describe a needed piece of equipment or service and specify what it must do. For example, the famous World War II jeep was a product of a design competition in which designers were given specifications in terms of top speed, climbing and carrying capacity, weight, cost, and so on. Many designers are invited to create prototypes, which are then subjected to rigorous independent tests. The designers have a more or less exact idea of what these tests will be, so they perform similar tests on their own prototypes before submitting them. Often the prototype-development process itself is funded by the government with an expectation that there will be a winnowing out of less promising designs over time. Ultimately, one or more designs are found to be particularly successful and are adopted on a large scale.

A school-reform model is not a jeep. Such models must be adapted to local circumstances, and there is an important benefit in having many successful models designed to accomplish similar goals so that educators can make informed choices among them. Yet in some respects, educational R&D could and should emulate the process that led to the jeep.

Imagine, for example, that the U.S. Department of Education were to announce a design competition to create highly effective and replicable methods and materials for the teaching of, say, beginning reading. The call for proposals could specify what kinds of outcomes would be assessed, what the limits of cost might be, and such other requirements as the use of stories and pictures that represent individuals of different ethnicities and genders in a fair, positive, and nonstereotypical way. Proposals would be invited to design, pilot, and formatively evaluate such programs. Say that the 10 most promising designs were initially chosen. Design teams might be funded for a period of time to do their initial development. This first set of design teams would be selected based on the quality of their ideas and experience, but also to represent a diversity of approaches. There would be an expectation that after the initial period the number of design teams going to the next stage might be reduced if some designs were not making adequate progress.

In a second funding period, design teams might be funded to pilot and formatively evaluate the outcomes of designs in comparison with control groups on measures linked to national or state standards. Finally, designs that appeared successful in their own evaluations would be evaluated by third-party contractors who would conduct relatively large-scale evaluations of the new programs in comparison with matched or randomly assigned control groups on measures selected by panels of experts to measure the broadest possible conception of high performance.

Of course, many programs exist today, and some of these already have their own evidence of effectiveness (see, for example, Fashola & Slavin, 1996). The FINDBEST system, currently being put into place by the U.S. Department of Education to collect evaluation evidence and consumer reviews of existing programs, would be a logical part of a process of identifying programs ready for third-party evaluation.

The design competition process I am proposing is similar to one implemented by the New American Schools (NAS). New American Schools, a private foundation funded by large corporate and foundation donors, held a design competition in 1991 in which applicants were asked to submit "break the mold" designs for schools able to bring all students to 21st-century standards. The request for proposals was fairly specific in specifying a focus on national goals; a comprehensive approach including curriculum, instruction, technology, assessment, parent involvement, and integration of school programs with community services; and an ultimate cost comparable to current costs. NAS received almost 700 proposals and initially funded 11 for a one-year development period. Nine of the programs were then funded for an additional two-year period for piloting their designs, and then seven were supported for a final two-year dissemination period. (As of this writing, we are in the final year of dissemination funding.) NAS is not funding an outcome evaluation comparing designs to control schools, but is supporting the RAND Corporation to do observations and process evaluations throughout the piloting and replication phases (see Bodilly, 1995). RAND is also currently collecting routine district-collected test data to provide some indicators of program outcomes. Several of the design teams, including ours, are carrying out their own outcome evaluations in comparison with control schools. NAS is concentrating on dissemination efforts in nine jurisdictions (Memphis, Dade County (FL), Cincinnati, Philadelphia, Pittsburgh, Seattle and four of its suburbs, San Diego, and the states of Maryland and Kentucky) and, in collaboration with the Education Commission of the States and the National Alliance for Restructuring Education, is helping these jurisdictions with systemic reforms intended to increase the chance that the school designs will work on a large scale and cause the jurisdiction to institutionalize a process of change. The jurisdictions were themselves selected in a competitive-proposal process; they had to commit themselves to implementing NAS or comparable designs in at least 30% of their schools by the year 2000. NAS is providing very limited funds to the jurisdictions. Instead, jurisdictions are paying the costs of implementing the NAS designs out of their own resources, such as Title I and reallocated state, local, and federal funds.

The ultimate success of the NAS approach is not yet assured, but it seems likely. As of fall 1996, Memphis has more than 30 schools implementing a total of six different designs, and Dade County has more than 50 schools implementing three designs. Maryland, the Seattle consortium, and Cincinnati are all moving forward with implementations of several designs. Collectively, there are about 400 schools implementing NAS designs. Many of the NAS designs, including ours (Slavin et al., 1996), are already producing convincing evidence of effectiveness from their pilot sites (see Stringfield, Ross, & Smith, 1996).

Whatever its ultimate outcome, the New American Schools experience has already demonstrated several important principles. First, it has shown that there is no shortage of designers willing and able to build new visions of education, to implement them, and to disseminate them. The NAS designs are very different from each other, but all are well thought out, comprehensive, and compelling. Some have proven more popular than others with schools, and some have developed more capacity to work with large numbers of schools than others, but all of the remaining seven designs are working in multiple locations beyond their pilot sites. Second, it is apparent that even in a time of diminishing discretionary resources, schools and districts are eager to adopt proven or promising programs tied to national standards. The NAS jurisdictions have generally made some funds available to schools to help them with start-up costs, but the overwhelming majority of funds to support design adoptions come from the schools' own resources. Third, NAS has shown that funders—in this case, primarily large corporations and foundations—are willing to provide substantial resources to fund educational R&D if they see a good chance that it will result in concrete, measurable change. The money raised by NAS, more than $120 million over a five-year period, is not much less than what OERI spends on all of its research centers over a similar time period. (Current funding for the 10 OERI research centers is about $28 million annually.) It seems likely that members of Congress and other policymakers would also be far more likely to appropriate large sums of money for educational R&D if they saw a direct relationship between their investments and widespread adoption of practical approaches to curriculum, instruction, and school organization of known effectiveness on assessments linked to national standards. There would still be a need for the more basic and informational research that is currently the focus of OERI-funded centers and field-initiated research; program adoption should by no means be the only route to reform. If anything, an emphasis on design competitions should enhance the status and importance of more basic research, just as basic research in biology is given great importance by its obvious (if often indirect) contributions to progress in medicine. A new drug or medical procedure requires enormous investments in applied research in preparation for a rigorous FDA approval process, but builds on basic research. A similar relationship between basic and applied research in education, leading to products of great benefit to children, could give all of education R&D the status and funding it deserves.

Elements of a Design Competition System

To have the desired outcomes, design competitions would have to exist within a supportive larger structure. As always, the details matter. The following sections discuss some of the key issues a design competition would have to address.

Selection of Development Priorities

The entire design competition process would be expensive and would engage the productive energies of many of our best developers and researchers. It would be critical to ensure that the objectives addressed by each are of the greatest possible importance to the practice of education, so obviously important and noncontroversial that they would be universally supported. In addition to this, design competitions should be focused on issues that seem capable of convincing resolution with evidence.

A few examples of issues that seem appropriate for design competitions are as follows:

Beginning reading methods. There are widely divergent theories of how best to teach reading in grades K–2, but little rigorous field research to validate these theories. Success in reading at the early grades is perhaps the most important predictor of school success that is directly under the control of schools. This seems to be an obvious choice for a first application of a design competition process.

Early mathematics methods (K–3). Like early reading, early mathematics is of great importance, and there are many quite different ideas of how to teach it. Objective, relatively unassailable evidence of the effectiveness of alternative approaches would be of enormous importance and is capable of being produced.

Math and science programs. It is in math and science that the U.S. trails many of its economic competitors, and math and science instruction in upper elementary, middle, and senior high school would be critical foci for design competitions. The National Science Foundation and other federal agencies have funded an enormous amount of development in these areas in recent years, but there is little objective evidence of the effectiveness and replicability of these new curricula. Third-party evaluations in this area could probably start immediately and are critical.

Middle school organization. Many children succeed at some level throughout elementary school but then run into serious difficulties in middle school. For 20 years, various versions of "middle school models" have been advocated, implemented, and debated, but rarely evaluated. In particular, there is a need for well-developed models that minimize ability grouping, but it is important to assess all types of approaches, regardless of philosophical or theoretical bases.

School-to-work programs. There are many alternative approaches attempting to infuse content related to careers and technical education into the secondary curriculum and to improve school-to-work linkages for students not going to four-year colleges.

High school completion and college entrance programs. There are several quite promising approaches to ensuring that all children take courses that lead to college and pass gatekeeper courses (e.g., algebra), such as Equity 2000 (Jones, 1993). There are also programs, such as Project AVID (Swanson, Mehan, & Hubbard, 1995), that help ensure that talented poor and minority children do what is necessary to go to college. Programs of this type are in need of further development, evaluation, and dissemination.

Of course, there are many other priorities that might be established for design competitions, such as writing, social studies, or foreign language programs; bilingual education models; early childhood programs; classroom management approaches; special education/inclusion models; computer literacy; and so on. Additional comprehensive methods, such as our own Roots and Wings program and other New American Schools-funded designs, are particularly important to develop and evaluate, and there is an immediate need for rigorous third-party evaluation of the existing NAS models and other schoolwide designs. However, we cannot start with everything at once, so it is reasonable to begin with objectives that clearly lie on a critical path from school entry to successful transition into college or the workforce for the largest possible number of students, especially those placed at risk of school failure. Squeamishness about setting priorities cannot be allowed to hold back a process likely to solve some of our most important education problems. Development priorities, specifications of design requirements, and selections of measures could be done by broadly representative panels of experts on the content, assessment, research design, and professional development. (More on this later.) Additional priorities could be added over time.

Solicitation and Selection of Design Teams

Once a design priority and specifications were established, funds would be set aside both for funding the development process and for third-party evaluations. Design teams would be asked to submit proposals responsive to the design specifications. The solicitation of design teams should be open to a broad range of groups, public and private, university-based or not.

Design teams would be allowed to enter the design competition at any point on the development/evaluation continuum, depending on their current status. For example, some teams might propose several years of development before moving to formative evaluation. Others (such as existing NDN or NAS designs) might ask for funds to complete or refine a previously developed model and move quickly to formative evaluation. Others might have already done extensive formative evaluation with convincing evidence of effectiveness and may apply immediately for third-party evaluation. As noted earlier, the U.S. Department of Education's FINDBEST system might help identify programs ready for third-party evaluation. In other words, the design competition process should not assume that all designs are starting from scratch and should not hold all design teams to the same schedule. The process should not head toward one massive "horse race" evaluation comparing all innovative models (like the Planned Variation Follow Through evaluation of the early 1970s; see Rhine, 1981), but rather should head toward comparisons of designs to current widespread practice whenever the designs are ready (as is done by the Food and Drug Administration). A "horse race" gives the false impression that there is one best model, and perhaps more importantly, holds all models to the same schedule, evaluating some that may be too early in their own development/evaluation process (as happened in Planned Variation). Design teams should be added over time to contribute to an ongoing design and evaluation process encompassing many design teams at different stages of development, with third- party evaluations beginning only when designs are ready.

Over time, funding for some designs might be discontinued. This would happen if the designs did not seem to be progressing well, if the replicability of the designs seemed unlikely, if designs were unsuccessful in formative evaluations, and so on. The designs should not be actively placed in competition with each other (by, for example, announcing in advance that half will be discontinued) because cooperation among design teams should be encouraged. However, there must be a mechanism for reviewing progress and abandoning clearly unpromising designs.

Formative Evaluations

Design teams would be funded to conduct their own formative evaluations, patterned closely on the summative third-party evaluations (see below). Designs could not go to third-party evaluation unless they presented their own convincing evidence of effectiveness on outcomes similar or identical to those used in the third-party evaluation.

Third-Party Evaluations

The linchpin of the entire design competition is rigorous, objective, broad-based third-party evaluations of promising designs. Anything less would deprive the design competition system of the external verification it must have to maintain its credibility with educators, policymakers, and the general public. The promise (or threat) of third-party evaluation would ensure the integrity of the entire system, from design to formative evaluation; designers would be unlikely to create designs that are merely attractive, easy to implement, or in line with current fads. They must create designs that will help children achieve to high standards.

One key issue in third-party evaluations is the nature of the outcome indicators used. Previous third-party evaluations, such as Planned Variation, were criticized for focusing on narrow objectives that could be assessed on standardized measures. To avoid this type of criticism, evaluations would have to assess a very broad range of possible outcomes. For example, a third-party evaluation of beginning reading models might include measures of oral and silent reading, word attack, comprehension, and other traditional outcomes, as well as such outcomes as the number of books read independently, attitude toward reading, ability to predict outcomes of a narrative, and perhaps creative writing. As a practical matter, matrix sampling (different students take different portions of a comprehensive assessment) could be used to obtain the broadest possible assessment of reading without overburdening each child. A likely outcome of very broad measurement of this kind would be that some programs would do well on some measures, others on different ones. This is fine. Educators could then decide which outcomes are most important to them in choosing among programs.

The specific measures to be used in third-party evaluation would be selected by a panel of experts, probably from among existing assessments (such as state performance assessments) closely linked to state or national goals. Care would have to be taken to be sure that developers could not "teach to the test" in any narrow way. For example, they might know that their students would be tested on particular domains of knowledge and skills but not see individual items in advance.

Third-party evaluations should compare schools using innovative programs to schools using "traditional methods." The definition of what is "traditional" is critical here; researchers would have to establish what practices were currently in widespread use and see that control schools were using some form of these methods.

Ideally, schools for the third-party evaluations would be chosen at random from among schools that volunteered to use the program being evaluated. For example, schools in a given district might be asked to volunteer to implement a new middle school model. This offer might be made in 5 to 10 districts around the country: some urban, some suburban, some rural, some with language-minority students, some large schools, some small ones, and so on. Fifty schools might be identified. Twenty-five might be randomly assigned to use the program and 25 to serve as controls (and to implement their current programs for a few more years). Control schools would receive extra resources, partly to balance those given to the experimental schools and partly to maintain a level of motivation to serve as control groups.

In addition to assessing student outcomes each year, evaluators would conduct process observations and careful cost estimates to characterize what was done and how school practices changed over time.

The third-party evaluators would be contractors selected for their strict neutrality with respect to the innovations being evaluated. Every effort would be exerted to see that political or other considerations had no impact on their evaluations or their reports. The evaluators would not be asked to provide a "thumbs up" or "thumbs down" for the entire design; instead, they would be asked to identify where the designs were working and where they were not and to attempt to find connections between particular program practices and program outcomes. For example, they might conduct process/product studies to correlate implementation of various program outcomes with program outcomes (in control as well as experimental schools) and to relate descriptive or ethnographic data to program outcomes. Evaluators would report outcomes for various subgroups of students or schools as well as overall outcomes.

Dissemination of Effective Designs

Successfully evaluated designs should be very popular. Supported by rigorous evaluation evidence and given a meaningful stamp of approval by the federal government, they should be attractive to schools and districts. Publishers and other for-profit organizations would probably be eager to disseminate the designs. However, there may be a need for government to provide dissemination funds to help design teams that are in nonprofit institutions and wish to remain there.

A large and growing shelf full of truly proven, replicable models should have a profound impact on educational policies at many levels. For example, if such a shelf existed, Title I and state compensatory education funds might shift away from pull-out teachers and aides, long known to be ineffective (Puma, Jones, Rock, & Fernandez, 1993; Slavin, 1994), toward helping high-poverty schools adopt proven models. Professional development programs, such as Eisenhower and Title VI, might support program adoptions.

It would be a mistake to *require* adoption of proven programs. Educators should be able to make informed choices in how they intend to bring their students to national standards. Because the development/evaluation process would take many years, schools might wish to choose models that are at some point in the process but are not yet evaluated by third parties. For example, programs that had successfully completed their own formative evaluations would already be far better validated than any but a handful of today's methods. However, low-performing schools failing to select proven models might be asked to justify their choice and to collect their own data on whatever alternative they choose, especially if they are using federal funds (such as Title I) or if they are schools that are losing ground on valid state assessments.

The potential misuse of evaluation data must be guarded against. One particular problem to be concerned about would-be developers who provide lavish amounts of professional development, monitoring, and other services or materials to schools involved in the formative and summative evaluations, but then provide much less to schools implementing the program after it has been validated in third-party evaluations. This possibly could be minimized by having third-party evaluators carefully document the services provided to schools and then having occasional reviews of validated programs (including random site visits) to ensure that program quality and intensity are being maintained. Publishing contracts and other agreements might be reviewed to see that there are guarantees that schools will receive the amounts and kinds of services found in the evaluation to be critical. This is a key point, as publishers often seek to sell materials (which are usually profitable) and minimize professional development (which is rarely profitable).

Clearly, any dissemination of proven practices would have to focus on the quality of implementation. While it is not true that any program will work if it is well implemented, it is true that poor implementations can undermine the effectiveness of any program, no matter how well validated. There is a tendency in American education for schools to want to appear to be implementing various innovations without actually having to make important changes in their practices. To the extent that government funds the dissemination and/or adoption of proven programs, it must seek assurances that program implementations are of high quality and provide adequate amounts of professional development and follow-up.

Can We Afford Design Competitions?

A design competition system would be expensive. As noted earlier, New American Schools will have spent about $120 million in five years, or $24 million per year, to fund the development of seven whole-school designs. It is true that NAS has spent a great deal of these funds on dissemination and on systemic-reform issues, but at the same time, it has not invested as much in third-party evaluation as I am recommending. For the sake of argument, I'll use this figure as an estimated cost for each design competition. By standards of current funding for education R&D, $24 million is a lot of money—almost equal to the funding ($28 million) for all of OERI's 10 research centers, for example. If multiple design competitions were under way simultaneously, as should be the case, the cost could easily consume all of OERI's current budget.

However, in the broader context of education funding, the cost of design competitions would be a trivial proportion. For example, even after recent cuts, Title I funding is about $6 billion. If 1% of this were set aside for design competitions, this would produce $60 million per year, enough for two or three design competitions (see Commission on Chapter I, 1992, for a similar recommendation). Traditional uses of Title I/Chapter 1 funds have never been found to have large impacts (Puma et al., 1993; Carter, 1984). Imagine that this 1% were invested in design competitions and third-party evaluations of existing and future programs in early reading, early math, and middle school design. This would be certain to produce better methods and better knowledge about these methods for use in Title I programs. Could anyone possibly argue that we cannot afford to spend 1% of Title I funding to make the other 99% produce positive results for children? Yet Title I is itself a tiny proportion of all education costs.

There is a case to be made that the federal government is too cumbersome, political, and regulation-driven to do a good job of managing design competitions. In fact, it may be difficult for a government agency to have the flexibility and freedom to, for example, drop support for programs that are not headed toward success. It might be better to have federal funds for this purpose managed by a private nonprofit organization or other independent group.

Perhaps it is quixotic to propose vast new expenditures in educational R&D at a time when we are struggling to maintain the minimal investment we have now. Perhaps we should be looking once again to the foundations and corporations that have provided the critical support behind almost all current reform networks. Yet I believe that we are entering an age of accountability in government, in which government programs must prove their effectiveness to maintain or expand their funding. In an age of "prove it or lose it," we cannot continue to hope that political support alone will continue indefinitely to justify federal investments in education. Educational researchers must make a case to advocates for Title I and other federal education programs that the interests of children, not to mention the narrowest interests of schools in maintaining federal education funding, depend on learning how to turn money into achievement. Design competitions are at least one means of moving in this direction.

Conclusion

Adding a system of design competitions would be a radical shift for educational R&D. Yet it is an essential one. Not until the education research community is able to routinely produce solid, unassailable advances in educational practice will the entire R&D enterprise be taken seriously, be funded adequately, and make a major difference in the lives of children.

Notes

This paper was written under funding of the Office of Educational Research and Improvement, U.S. Department of Education (Grant No. OERI-R-117-D40005). However, any opinions expressed are those of the author and do not necessarily represent the positions or policies of the U.S. Department of Education.

I would like to thank John Anderson, Liz Berry, Kalman Hettleman, Nancy Madden, and Tom Glennan for their comments on an earlier draft of this paper.

References

Bodilly, S. J. (1995). *Lessons from New American Schools Development Corporation's development phase*. Washington, DC: RAND.

Carter, L. F. (1984). The sustaining effects study of compensatory and elementary education. *Education Researcher, 13*(7), 4–13.

Comer, J. (1988). Educating poor minority children. *Scientific American, 259*, 42–48.

Commission on Chapter 1. (1992). *Making schools work for children in poverty*. Washington, DC: Council of Chief State School Officers.

Corcoran, T., & Goertz, M. (1995). Instructional capacity and high performance schools. *Educational Researcher, 24*(9), 27–31.

Fashola, O. S., & Slavin, R. E. (1996). *Effective and replicable programs for students placed at risk in elementary and middle schools*. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed At Risk.

Goertz, M. E., Floden, R. E., & O'Day, J. (1996). *Studies of education reform: systemic reform*. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Jones, V. (1993). *Equity 2000: Commitment to excellence and equity*. New York: The College Board.

Levin, H. M. (1987). Accelerated schools for disadvantaged students. *Educational Leadership, 44*(6), 19–21.

Pinnell, G. S., DeFord, D. E., & Lyons, C. A. (1988). *Reading Recovery: Early intervention for at-risk first graders*. Arlington, VA: Educational Research Service.

Puma, M. J., Jones, C. C., Rock, D., & Fernandez, R. (1993). *Prospects: The congressionally mandated study of educational growth and opportunity. Interim report*. Bethesda, MD: Abt Associates.

Rhine, W. R. (1981). *Making schools more effective: New directions from follow through*. New York: Academic Press.

Sizer, T. (1984). *Horace's compromise: The dilemma of the American high school*. Boston: Houghton Mifflin.

Slavin, R. E. (1989). PET and the pendulum: Faddism in education and how to stop it. *Phi Delta Kappan, 70*, 752–758.

Slavin, R. E. (1994). School and classroom organization in beginning reading: Class size, aides, and instructional grouping. In R. E. Slavin, N. L. Karweit, B. A. Wasik, & N. A. Madden (Eds.), *Preventing early school failure: Research on effective strategies* (pp. 122–142). Boston: Allyn and Bacon.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1994). Roots and Wings: Inspiring academic excellence. *Educational Leadership, 52*(3) 10–13.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Every child, every school: Success for All*. Newberry Park, CA: Corwin.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). "Whenever and wherever we choose . . . ": The replication of Success for All. *Phi Delta Kappan, 75*(8), 639–647.

Smith, M., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233–267). Bristol, PA: Falmer.

Stringfield, S., & Herman, R. (1995, April). *The effects of promising programs on students: Results from the Special Strategies Study*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Stringfield, S., Ross, S., & Smith, L. (1996). *Bold plans for school restructuring: The New American Schools designs*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Swanson, M. C., Mehan, H., & Hubbard, L. (1995). The AVID classroom: Academic and social support for low-achieving students. In J. Oakes & K. H. Quartz (Eds.), *Creating new educational communities* (pp. 53–69). Chicago: University of Chicago Press.

Vinovskis, M. A. (1993). *Analysis of the quality of research and development at the OERI research and development centers and at the OERI regional educational laboratories*. Washington, DC: U.S. Department of Education.

Y

RESEARCH NEWS AND COMMENT

Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop, and Share Promising and Exemplary Products and Programs

SUSAN S. KLEIN

Thanks to Robert Slavin for "Design Competitions: A Proposal for a New Federal Role in Educational Research and Development" (1997a). This response is intended to show how the U.S. Department of Education (ED) and its Office of Educational Research and Improvement (OERI) are starting to develop a consumer-oriented evaluation system that is based on many of the same premises as the design competitions and that would complement and optimize the substantial federal investment needed for each design competition. In brief, Robert Slavin proposed that the federal government fund the development, refinement, and evaluation of a winning group of instructional programs that are most likely to meet predetermined specifications related to "significantly raising student achievement" (p. 22). In his proposal, he also mentioned that "the U.S. Department of Education's FINDBEST system might help identify programs ready for third-party evaluation" (p. 25). This article discusses these and many other similarities and differences in the premises and potential strategies for these two approaches to provide more systematic leadership in federal education development, evaluation, and dissemination. It also will describe some key aspects of the evolving System of Expert Panels including a name change from its initial working title of the Findbest System.[1]

The System of Expert Panels can be viewed as an emerging consumer-oriented evaluation component of the National Education Dissemination System (NEDS).[2] The development of this system is already benefitting from the New American Schools Design Competition as well as from ideas in Dr. Slavin's proposal. After providing an overview of the evolving System of Expert Panels, I will discuss similarities and differences in these two approaches to improve federal R&D management and conclude with recommendations on how to coordinate them.

Overview of the System of Expert Panels

Background and Purposes

Under its 1994 reauthorization (Title IX of Goals 2000: Educate America Act), OERI has been given responsibility for

- Developing standards to designate promising and exemplary products, programs, and practices (draft regulations for these standards published June 3, 1996, were revised in response to public comment and will be published by summer 1997;
- Establishing a System of Expert Panels to make recommendations on these designations to the secretary of education (two pilot panels in the areas of mathematics and science and gender equity were appointed in 1996); and
- Coordinating dissemination activities and programs not just within OERI, but with other parts of the department, with other agencies concerned with education, and with associations and other levels of government, such as state education agencies and local school districts.

The System of Expert Panels should enable the federal government to

- Help the public learn about the comparative advantages of what exists among the many available replicable R&D-based products, programs, practices, and policies[3] based on an understanding of their relative merits[4] and
- Help federal and other funders and producers of R&D-based resources maximize their investments by providing support for further evaluation, improvement, and dissemination of existing promising and exemplary R&D-based resources and by identifying gaps that may help indicate areas where new products and programs are needed.

Other expected benefits include providing practical reasons for educators (and, in some cases, students) to improve their evaluation skills related to the use of evaluation information in making decisions on instructional tools such as programs or products. Participants in system activities will also be able to contribute to the development of new methods to combine evaluation and dissemination functions to better serve education consumers.

This system is congruent with the U.S. Department of Education's strategic plan (1994) to better serve its customers, but it is not limited to reviewing models that have been developed with previous federal funding. It builds on, but differs from, the Joint Dissemination Review Panel and its successor, the Program Effectiveness Panel, used by the department's National Diffusion Network (NDN), which is no longer funded.[5]

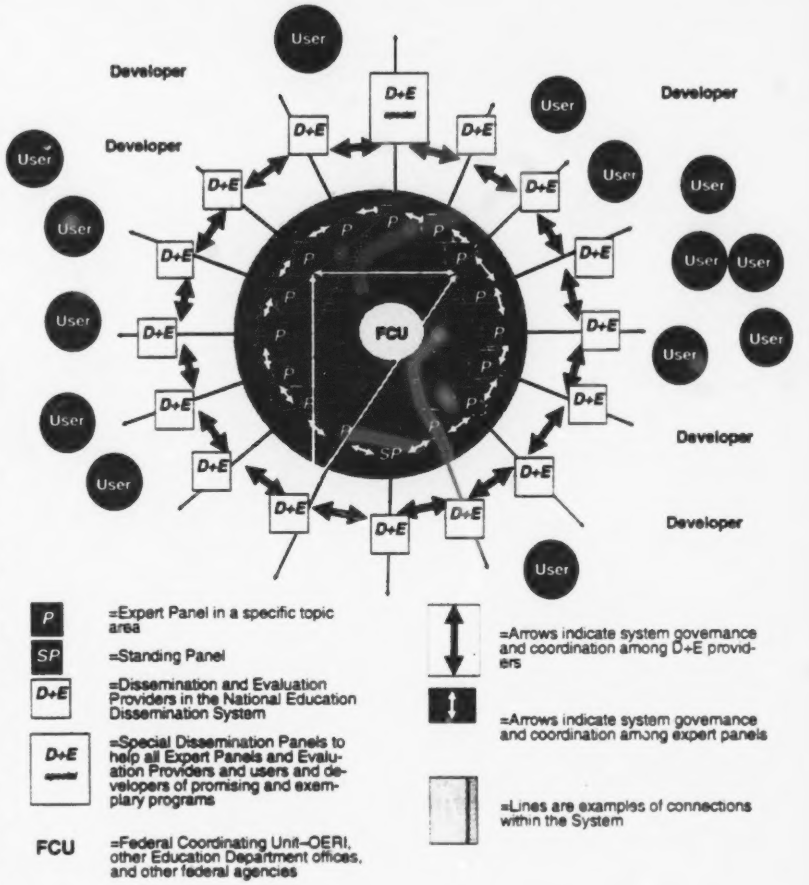

User
Developer
D+E
P
SP
FCU
P =Expert Panel in a specific topic area
SP =Standing Panel
D+E =Dissemination and Evaluation Providers in the National Education Dissemination System
D+E =Special Dissemination Panels to help all Expert Panels and Evaluation Providers and users and developers of promising and exemplary programs
FCU =Federal Coordinating Unit–OERI, other Education Department offices, and other federal agencies
=Arrows indicate system governance and coordination among D+E providers
=Arrows indicate system governance and coordination among expert panels
=Lines are examples of connections within the System

FIGURE 1. *Wheel diagram of the key elements of the system of expert panels.*

An Image of This System That Includes More Than Expert Panels

It is important to develop a well-understood and articulated nationwide system that has clear connections to other complementary systems, such as the National Education Dissemination System (NEDS) and the even more loosely configured R&D production system. Ideally, this system would be logical, inclusive, and dynamic. As explained later, the federal role would be facilitative and collaborative, rather than controlling.

As currently envisioned, the System of Expert Panels would be more than a collection of individual expert panels—a justification for choosing a better name. Figure 1 shows that the system would be led and coordinated from the *inside hub circle* by a federal coordination unit (FCU) of experts from various U.S. Department of Education offices and other agencies involved in education activities to identify and share the best from R&D. This FCU would include agency representatives such as federal staff liaisons with current or potential expert panels.

The *middle circle* connects the individual expert panels (P) in a comprehensive array of topic areas. Where feasible, each topic-focused expert panel would have liaisons from federal offices who would be in a position to use money and knowledge from their programs to contribute to and benefit from panel work. These federal liaisons would also advise on the selection of panel members. This circle also contains a standing panel and a democratic governance and coordination structure for the semi-autonomous expert panels that would represent their topic-focused constituency groups.

The *outside circle* represents the consumer-oriented dissemination and evaluation providers (D+E). But, unlike the other parts of the diagram, this outer circle would be composed primarily of existing dissemination and evaluation entities such as ERIC, technical assistance centers, regional educational laboratories, organizations that provide third-party evaluations, national R&D centers, publishers, participants in design competitions, and organizations managing ED and other related World Wide Web sites. Some of these D+E providers would specialize in specific topic areas and some in multiple topic areas. In some cases, national topic-specific dissemination and evaluation organizations would be prime candidates to serve as support contractors for an expert panel in their topic area. For example, the Women's Educational Equity Act (WEEA) Equity Resource Center is the support contractor for the Gender Equity Expert Panel. If needed, new topic-specific dissemination and evaluation services could be created to support the work of expert panels in specific topic areas. It is likely that most D+E providers would also have key roles in other aspects of the NEDS and the more loosely configured R&D production system. Like the middle circle, participants in this outer circle would be linked across panel topics by a governance structure[6] and by joint use of special D+E providers. Functions of topic-specific and multi-topic special D+E providers would include disseminating promising and exemplary programs, obtaining and sharing systematic user feedback, preparing reviews and evaluation reports on promising and exemplary programs in multiple topic areas, and sharing information about opportunities to participate in the third-party evaluations of promising programs. In a commissioned paper, Patricia Campbell (1994) described initial ideas on how a D+E provider could obtain, synthesize, and share teacher and student evaluation feedback on promising and exemplary programs.

The spokes of the wheel show that much of the coordination of the work and contact with the users would be organized in relation to each of the topic-focused expert panels. However, there would also be many other connections because some dissemination and evaluation structures cover multiple topic areas and because promising and exemplary programs (i.e., gender equity in mathematics, science, and technology) may be appropriately disseminated under the auspices of more than one topic area. (See Klein, 1997, for a more detailed discussion of these elements of the System of Expert Panels.)

Facilitative Federal Leadership Role

The federal government will play a facilitative leadership, partnership, and community-building role more than a regulatory role. This new type of collaborative leadership with the field can be developed by establishing expert panels in a wide variety of topic or special interest areas and sustaining them over many years with relatively modest federal investments because the topic-focused constituency groups would volunteer reviewing and other services, much as they do now in reviewing journal articles. In addition to strengthening support for the agency from its concerned constituencies, this approach should help these topic-focused communities develop a systematic process to share what works and to identify gaps. The current panels are succeeding in this in two different ways. The Gender Equity Expert Panel has created an advisory group of over 100 experts who are willing to help the expert panel. And many of the members of the Mathematics and Science Expert Panel are leaders of key constituency groups in their area. This closer connection with constituency groups who have expertise in particular topic areas should also help make agency-supported work in areas such as design competitions and technical assistance activities more relevant to constituent needs.[7]

A facilitative federal coordination function is needed to help all components of the system learn from each other and work in complementary ways where there are natural overlaps in responsibilities. More specifically, it will

- Provide leadership in coordinating (and, where feasible, funding) all components of the wheel diagram;
- Develop systematic evaluation and dissemination methodologies and consistent decision rules and definitions for system activities so that the public will understand distinctions between "promising" and "exemplary" programs; and
- Develop incentives to make the system work.

Coordination of components of the wheel diagram requires flexibility to take advantage of the interests of various federal offices and constituency groups. In doing so, it is assumed that federal programs inside and outside of OERI would participate in this system, but that each would have somewhat unique activities to take advantage of their own legislative responsibilities. For example, in mathematics and science education, the National Science Foundation might be able to support a D+E provider to manage comparative third-party evaluations of promising programs with similar purposes and to design and manage design competitions to fill identified gaps, and ED might be able to support state professional development programs that choose promising and exemplary products and programs recommended by the Mathematics and Science Expert Panel or the Gender Equity Expert Panel.

Development of common evaluation and dissemination methods and definitions has started. The draft OERI standards for designating promising and exemplary programs have common criteria categories (evidence of effectiveness/success, quality, educational significance, and usefulness to others) and decision rules. (See Figure 2).

Development of major system incentives revolves around consequences for receiving designations of "promising" or "exemplary." It is suggested that if a program is designated as promising, it will merit additional investment for improving it and evaluating it so that it could justify a future exemplary designation. Based on this understanding, a promising designation would encourage only cautious adaptation or adoption, during which time it would undergo careful evaluation. Users would be encouraged to select exemplary rather than promising programs unless no exemplary programs meet their need or unless they wish to participate in the field testing of promising programs. The key purpose for designating programs as exemplary is to increase the use of what works best. In addition to developing working relationships with specific D+E providers, the expert panels would work with others to publicize information on exemplary programs through the World Wide Web, TV, and print media to help consumers choose from a variety of particularly worthwhile options.

The active partnership roles for the topic-focused constituency groups will need to build on their strengths, resources, and interests. For example, the Gender Equity Expert Panel has formed six subpanels to better connect with specific interest groups, and the Mathematics and Science Expert Panel members have many contacts with mathematics and science educators they can tap through state affiliates of their national associations.

Consumer-Oriented Evaluation

This focus should be particularly appealing to educators who know they can do a better job if they are able to choose and use effective tools or instructional programs. This concept focuses on obtaining and disseminating descriptive and evaluative information designed to help consumers make decisions about what products or programs will be best for them. In addition to educators, "consumers" include the entire public (ranging from policymakers to students) who might be interested in learning about the merit of an education product, program, or practice. Because many of these R&D-based tools have similar purposes, consumers will want to know their comparative merit on a range of criteria that are likely to be important for their own decision-making. The OERI standards categories as described in Figure 2 are intended to provide a framework for topic-focused criteria and for research-based criteria selected by experts. Ideally, this System of Expert Panels should use D+E providers to develop consumer reports[8] for specific topic areas in which there are expert panels. These consumer reports could be supplemented by descriptions and summary reviews of individual promising and exemplary programs that a potential user could retrieve via a computer/Internet search. Consumers should be able to use either or both of these information sources so they can do their own side-by-side comparisons and make informed decisions. Based on past experience with Educational Products Information Exchange (EPIE) (Komoski, 1989) and other sources of information, it is unlikely that education consumers will want to pay for this information. So it is assumed that the government, foundations, and associations will need to pay for D+E providers to do this public domain work. When all system components are operating, the evaluation results that would be included in this consumer information would not end with initial expert panel decisions. The panels and their support contractors would continue to update this information based on work of the D+E providers described in the outside wheel of the system diagram. For example, a special D+E provider would collect information from users of promising and exemplary programs, thus making sure that users play a continual role in helping each other learn what works for whom. It is assumed that this new feedback function and many of the nationwide operations of this system will become increasingly feasible as more educators use computers and the Internet to communicate.

This summary is based on February 11, 1997, draft regulations revised according to public comment and suggestions from the OERI board at its January 31, 1997, meeting in Washington, DC. A subsequent version of these regulations is to be published in the Federal Register *by the summer of 1997.*

A panel may recommend to the secretary that a program be designated as *promising* if the panel determines that the program is *strong* on each of the four categories of standards (evidence of success/effectiveness, quality of the program, educational significance, and usefulness to others). A panel may recommend to the secretary that a program be designated as *exemplary* if the panel determines that the program is *excellent* on each of the four categories of standards. The decision framework for determinations of *strong* and *excellent* on evidence of success/effectiveness is intended to be the same for all panels. However, for the other three standards categories—quality, significance, and usefulness to others—each panel may add to the core criteria and establish its own decision framework for designation as promising and exemplary. It is expected that the panels will require a *strong* judgment on almost all criteria within each of these three categories for a designation as promising and a judgment of excellence on almost all criteria within each of the three categories for a designation as exemplary. For a program to be designated exemplary, any weaknesses must be minimal and easily corrected.

(a) *Evidence of Effectiveness/Success.* To be judged *strong* in the category of evidence of effectiveness, all of the following criteria must be met. A program must

- **Have defensible overall evidence supporting claims of worthwhile performance results (without substantial harmful results) at one or more sites (without failing at a large number of other sites) and**
- **Have logical or other evidence of adaptability or transportability to other sites.**

The combination of this positive evidence of effectiveness and potential replicability creates a significant probability that the program will eventually be able to provide evidence to support claims of exemplary meritorious results as defined in the following section.

To be judged *excellent* under the category of evidence of success/effectiveness, all of the following criteria must be met by convincing evidence that

- **Very important claims of positive results (or performance outcomes) can be reasonably attributed to the program and**
- **Evidence to support these claims of worthwhile results was sustained in multiple site replications within the past few years (without failure at a large number of other sites or accompaniment by harmful results).**

(b) *Quality.* The panels will make their judgments about quality by reviewing the program materials and determining the extent to which the program

- **Is congruent with sound research and practice;**
- **Incorporates accurate and up-to-date information/content;**
- **Promotes equity and is free of bias based on race, gender, age, culture, ethnic origin, disability, or limited English proficiency;**
- **Is appropriate, engaging, and motivating for the intended audiences; and**
- **Contains materials that conform to accepted standards of technical product quality.**

(c) *Educational Significance.* The panels will use their expertise in the area to determine the extent to which the program

- **Addresses an important education issue, challenge, or problem; and**
- **Has advantages over other programs with similar purposes.**

(d) *Usefulness to Others.* The panels will make these judgments by using their knowledge of what is valued by educator and student users in determining the extent to which the program

- **Is reasonable in terms of costs to potential users in relation to expected benefits.**
- **Is or can be made easily available to potential users,**
- **Can be readily adopted or easily adapted in new locations, and**
- **Can be used in conjunction with other programs if appropriate.**

FIGURE 2. *OERI standards and criteria for expert panels to use to distinguish between promising and exemplary programs.*

Similarities and Differences in Premises and Strategies for the Proposed Design Competitions and the Evolving System of Expert Panels

In discussing these two proposed federal R&D management roles, it is important to remember that many of the elements of the System of Expert Panels are required in the OERI legislation. Design competitions, while compatible with development responsibilities in the legislation, are not specified. Also, the major pilot test for the design competitions with the New American Schools has been supported outside the federal government, and the pilot tests of expert panels in mathematics and science education and in gender equity are being supported by the U. S. Department of Education.

There are many shared premises, but often different strategies, to implement these two R&D management approaches. Shared premises include

- The need for a more deliberate and focused federal R&D management role;
- Strong commitment to using research, development, and evaluation to learn what works well to inform and foster revision (adaptation) and replication and thus improve education;
- A belief that national dissemination efforts should find multiple ways to encourage selection and use of the best education R&D has to offer; and

- A realization that there are many methodological, strategic, and resource challenges that will make either of these strategies difficult to implement fully and to sustain over the years. Key differences in strategies relate to (1) the federal role, (2) coverage of R&D outcomes, and (3) use of somewhat different evaluation approaches to determine the best.

Shared Premises

- **The need for a more deliberate and focused federal R&D management role.**

While both efforts acknowledge the importance of diverse types of federal investments in education research—ranging from field-initiated research to grant competitions to specific requests for proposals—they are designed to make federal investments in development, evaluation, and dissemination more purposeful, systematic, and, if merited, sustained. For example, many federal education programs fund "demonstration" projects that generally allow for the development of a creative program in one site. But when federal funding ends, the program may also end at the original site, and there are few opportunities to learn if it merits continued support for additional revision, evaluation, or dissemination to others. Similarly, federal offices rarely provide this type of additional support for meritorious programs not developed with federal funds. Thus, both design competitions and the System of Expert Panels are intended to carefully identify the gaps and opportunities and target subsequent federal funding to replicable programs with the greatest chance of helping students receive a better education and show improved performance. While most expert panels will be designed to cover the broad interests of their constituency groups and to last over a number of years, it is also possible for more short-term specific-focused expert panels to be established and funded by a federal office or for a broader panel to limit its search for solutions to priority areas during specific time periods (Datta & Scriven, 1997). This type of specification is similar to the focused approach of design competitions.

- **Strong commitment to using research, development, and evaluation to learn what works well to inform and foster revision (adaptation) and replication and thus improve education.**

Both systems want to help schools obtain good evidence on the effectiveness of school reform models so they will have justification for choosing among the models or sticking with their current practices. Both believe in prototype development (the "promising" designation in the standards for expert panels) and agree that only the best of these (the "exemplary" designation in the standards for expert panels) should receive continued support for widespread implementation. Neither has a rigid view of replication; both agree that careful adaptation makes sense and that the user must select not only the best things, but those that will complement other instructional approaches and fit the needs of users and the skills of staff.

Both assume that it is possible to find multiple effective options for various users and that some of the criteria that determine what's best are related to evidence that models have worked for others with similar populations, resources, and needs. Both have found that evidence of effectiveness is often lacking and assume that their approaches will encourage developers and others to collect better evidence and that the federal government may help pay for part of these evaluations.

The criteria would be similar for both. In addition to evidence of effectiveness to support important claims, both would pay attention to costs and to equity and cultural diversity. Both are likely to use some criteria based on national standards or professional consensus about research-based principles. Slavin's proposal states that "significantly raising student achievement on tests linked to national standards" is essential (1997, p. 22). The Mathematics and Science Education Expert Panel is planning to use national standards and benchmarks (Birman & O'Malley, 1996a, 1996b, 1997; Lacampagne, 1996), and the Gender Equity Expert Panel is using federal and state equity policies and shared professional understandings based on research syntheses to define acceptable indicators of gender equity because there are no specific national gender equity standards (Hanson, 1996). As with the specifications for a design competition, the expert panels will be expected to add many of their own criteria to the OERI standards for promising and exemplary programs. This is reflected in the submission guidelines for the Gender Equity Expert Panel (Maxwell, 1996) and in the extensive discussion of criteria by the Mathematics and Science Education Expert Panel (Birman & O'Malley, 1996a, 1996b, 1997).

Submissions to the System of Expert Panels and for design competitions are welcome during most of a program's development-evaluation continuum. There would be no need for a horse race to find the best model from those starting at the same time. Although design competitions are more likely to focus on the substitution of new things for common practice, they can easily support the continued development of promising alternatives as the New American Schools did for Robert Slavin's own work on Success for All and Roots and Wings. Both assume two important stages and types of evaluations. The initial evaluations for both would be somewhat formative, but they would also attend to performance outcomes, and they would probably be conducted by the developer. If a design team's evidence of success and judgments on other criteria are acceptable, the design competition managers would go to stage two and fund rigorous independent third-party evaluations. Similar third-party evaluation support would likely be made available for programs designated promising by an expert panel. Ideally, the third-party evaluators for both expert panels and design competitions would examine where programs are or are not working to find connections to outcomes and examine outcomes for subgroups of students.

- **A belief that national dissemination efforts should find multiple ways to encourage selection and use of the best education R&D has to offer.**

Neither is based only on recognition and awards for excellence. Both focus on learning about or developing what is likely to be useful to others. Both assume that most users will find the funds for implementation especially of exemplary models or designs in late stages of development. If programs are designated as exemplary by the System of Expert Panels or as long-term successes in a design competition, the federal government could encourage recipients of Title I or other federal funds to choose from among the exemplary programs or provide technical assistance or other support for their implementation. Incentives could also be provided to collect evaluation evidence on promising and exemplary programs to learn more about how they work in different situations. This evidence could be collected, analyzed, and shared by a D+E provider.

- **A realization that there are many methodological, strategic, and resource challenges that will make either of these strategies difficult to implement fully and to sustain over the years.**

Both will need to obtain major financial support, although because of the development work and built-in plans for extensive third-party evaluations of individual programs, the design competitions would be more expensive than an individual expert panel and the system-related work of its associated topic-focused dissemination and evaluation (D+E) providers.

Both may be opposed by educators who believe that replication of model programs and products is not a feasible strategy for educational improvement because of the need for adaptation or because they think educators must develop their own approaches based on their interpretations of research findings and their own insights and experiences.

Both realize that the technical and procedural challenges related to evaluating the designs of the participants in the design competitions or the programs to be examined by the expert panels are substantial. Examples include difficulty in agreeing on desired performance indicators, frequent lack of information on adequacy of implementation as well as performance outcomes, and practical difficulties in making sure that evidence collection and panel submissions will be "the whole truth and nothing but the truth." Both will also need to deal with the dissatisfaction of those who do not get advanced to the next stages in the design competition or are designated as "promising" or "exemplary" by the expert panels. Although both management strategies are intended to judge the program, not the developer, it is difficult to separate developers from their programs especially when the evaluations are likely to be dependent on the cooperation of the developers and thus not anonymous.

Also both could benefit from closer links to other systems (such as NEDS and R&D providers) so that these related systems would provide incentives and do other things to maximize the benefits and reduce the costs of the expert panel system or the design competition.

Key Differences in Strategies and Outcomes

Different federal roles. The System of Expert Panels is focused on supporting evaluation and dissemination, while the design competitions focus on funding strategic development. The System of Expert Panels is intended to change and systematize many aspects of federal support for development, evaluation, and dissemination to maximize the use of existing promising and exemplary models. For it to succeed, it will need to be well integrated with other federal activities. The design competitions are focused on fewer topics and probably would be limited to producing solutions in areas of high national need. Each design competition could be quite unique because they are independent of each other and don't even need to be managed by the same office.

In contrast, the more the System of Expert Panels grows and develops collaborative procedures both within and among expert panels and other system components, the more effective it should become. One of the benefits of the system is that it will help build or strengthen constituency groups around the what's-best issue. OERI is doing this by establishing a "Findbest listserv" advisory group for the System of Expert Panels and a "Gndrpan listserv" advisory group for the Gender Equity Expert Panel.[9]

Design competitions would require a major increase in federally funded education development, evaluation, and dissemination to create well-validated programs or practices and ensure their widespread use. Dr. Slavin proposes significant federal funding for a series of design competitions. His estimate of $120 million per competition over five years is based on experiences with the non-federally funded New American Schools, which had its initial design competition in 1991. Some new federal funding is also needed to support the System of Expert Panels, including third-party evaluations of promising programs, but it would not be as substantial as for design competitions because much of the system work would be shared with professional associations and volunteer advisors or it would piggyback on other federally funded D+E providers. This system would also use established dissemination mechanisms such as the Internet and federally supported clearinghouses and technical assistance providers.

This probably means that design competitions would only be feasible in areas where there is major funding. Also, the design competition managers may find it difficult to change support from an initial winner to something else if a better design option arrives. The System of Expert Panels would be designed to continually identify promising and exemplary models in a wide variety of areas under the leadership of topic-focused expert panels and their constituents. For example, the two pilot expert panels are in a well-funded area—mathematics and science education—and in a minimally funded area—gender equity in education. The lack of specific appropriations for the System of Expert Panels is both an advantage and disadvantage. The advantage is that many federal offices can participate to the extent their legislation allows as they manage their own education improvement responsibilities.

Coverage of R&D outcomes. While design competitions would focus on developing major new replicable models to meet challenges not generally well addressed by common practice, the System of Expert Panels is based on the assumption that many worthwhile products, programs, and practices have already been developed and emphasizes an active search for promising and exemplary models. Thus, each expert panel should yield many more replicable promising and exemplary programs than a design competition. This result is likely because programs reviewed by the expert panels may be smaller in scope and cover more topics. Smaller in scope means that practices, policies, products, or modules could be considered as well as a yearlong course or school-wide model. More topics would be covered because there would probably be more expert panels than design competitions and because multiple topics could be considered in the expert panels, especially when they establish subpanels with their own networks of experts. It is also likely that numerous programs will be judged promising because standards for "promising" would be quite reasonable and attainable. And finally, unlike design competitions, panel designations of promising or exemplary would not be limited because they are tied to commitments for future funding.

Use of somewhat different evaluation approaches to determine the best. The System of Expert Panels would focus its evaluations and reviews on facilitating consumer comparisons of promising and exemplary program options. The evaluation focus for the design competitions would be on improving the programs in the competition and on learning if they meet the competition specifications to solve the educational challenge better than common practice. Even though comparative evaluations among programs are not emphasized in the design competitions, Dr. Slavin points out one way this may be done in third-party evaluations. He suggests that information on many relevant performance outcomes should be described and that it is natural for some programs to do better on some outcomes than others and for educators to choose programs that do well on the outcomes that are most important for them. Because it should be easier to know the universe of what is best in a more limited topic area, expert panels, with assistance from their support contractors, should be able to make comparisons more easily than the more general review panels used by the National Diffusion Network (NDN).

While both management strategies value third-party independent evaluations for the more advanced stages of the design competition or for programs already designated promising, it is likely that expert panels would be more flexible in the evidence they review and accept to support claims of effectiveness and usefulness to others. For example, an expert panel might designate a program as exemplary if the program had convincing evidence from its own evaluation and there was some way to verify the accuracy and completeness of this evidence. However, expert panels would certainly appreciate receiving evidence supported by the more traditional third-party evaluations. Assuming they were excellent on criteria in all three criteria categories, programs from design teams that did well in these independent third-party evaluations would probably be judged exemplary by an expert panel.

The methodological challenges for evaluations in the System of Expert Panels are more extensive than for design competitions because less traditional evaluation approaches would be encouraged to obtain desired results. In addition to providing excellent information about the best available options, the system activities need to be designed to encourage wider reliance on and participation in evaluations of these replicable educational resources. A premise of the design competitions is that all who pass through the various design stages will automatically receive appropriate evaluation support, more formative in the early stages, and a rigorous third-party evaluation in the final stages. The System of Expert Panels will be in a position to try a variety of approaches to obtain good evaluation information. For example, the two pilot panels have said that they want to make sure that the submitters document their best evidence and thus plan to provide advice on revising submissions as necessary and to encourage resubmissions if new supporting evidence of effectiveness and utility can be obtained. It is also anticipated that federal agencies and foundations would support the continued evaluation of programs designated as promising and that, where feasible, these evaluations would be done in a comparative, but noncompetitive fashion using the same performance indicators for programs with similar purposes. The design competitions put less emphasis on comparisons among models with similar purposes. Rather, the third-party evaluation in the design competitions emphasizes comparison with common practice rather than encouraging competition among the teams. For example, Rand is now comparing student performance on the New American Schools models with student outcomes in the same district to make the case that the model is successful. To augment this information, the Rand evaluators will use some common indicators for finalists, but this is being done so that the different models don't feel like they are competing with each other. The third-party evaluations from the design team would also be able to look at more causal research questions relating to implementation than is likely for the potentially more numerous, but lower-cost, third-party comparative evaluations of promising or exemplary programs. Another challenge for the System of Expert Panels is to develop new ways to connect dissemination and evaluation incentives such as disseminating free copies of promising and exemplary products and technical assistance services to schools who deliver requested evaluation information.

Recommendations on Coordinating Design Competitions and the System of Expert Panels

Many in the educational R&D community welcome serious national attention to evidence of effectiveness, a key premise of both design competitions and the System of Expert Panels (Canine, 1995; Scriven, 1993, 1995; Seidman, 1997). Now we have a chance of making this focus a reality. Both efforts have benefitted from some pilot test experience, and both are supporting formative evaluations (Bodilly, 1996; Bodilly, Purnell, Ramsey, & Smith, 1995; Campbell & Bachmann, 1996). OERI planners of the System of Expert Panels have also sponsored meetings (Bogart, 1996; Schmieder, 1996) and commissioned papers and other research and analyses to learn from their own and others' experiences (Backer, 1995; Barkdoll, 1995; Campbell, 1994; Datta, 1994; Datta & Scriven, 1997; Hollifield, Stringfield, & Herman, 1996; Klein, 1993, 1995, 1996; Klein & Gwaltney, 1991; LaFollette, 1992; Muscara, 1996; Turnbull, 1996).

Because they are both at early stages and share many premises, future plans for each could be coordinated. For example, both could use similar criteria and decision rules like the OERI standards for "promising" and "exemplary." Also, as expert panels gain a full understanding of what is already working in their topic areas, they could be key advisors in developing design competition specifications for work needed to fill the gaps or address new opportunities. Similarly, if funds are secured for new design competitions in areas where no expert panels exist, an expert panel could be established to learn what exists that might contribute to the design goals before developing the design specifications. Such an analysis might lead to support for a comparative evaluation of promising and exemplary programs in the area as well as a request for funding something that has different purposes than these existing programs. One of the challenges faced by managers of design competitions is to review the work of design teams at various stages in their development to decide if they merit continued funding. An expert panel or a subpanel with the design competitions' specific purview could serve as an objective review body to determine which models in the design competition should be judged promising after their formative evaluation[10] or exemplary after receiving results from their third-party evaluation. Thus, coordinating design competitions with the System of Expert Panels will help the U.S. Department of Education make wiser choices about its investments in development, evaluation, and dissemination, and it will help the department assume a more active leadership role in working with its producer and consumer constituencies.

Notes

This article is intended to promote the exchange of ideas among researchers and policymakers. The views are those of the author and no official support by the U.S. Department of Education is intended or should be inferred. I would like to thank Dr. Slavin, other *ER* reviewers, Dr. Lois-ellin Datta, and Dr. Sharon Bobbitt, Director of the KAD, for comments on an earlier version of this article.

[1] Some suggest that we can do better than the initial or current working titles of "The Findbest System" or "The System of Expert Panels." Because one premise is that the experts in the R&D communities should assume a major partnership role in this system, please feel free to suggest a better name and share your ideas relating to this evolving effort by contacting the author: sue_klein@ed.gov, 202/219-2038.

[2] The OERI legislation says that its Office of Reform Assistance and Dissemination (ORAD) should provide leadership in coordinating federal and other dissemination programs and activities of the NEDS.

[3] These replicable R&D-based resources (1) can be used outside of the original development site, (2) are based on principles from educational research, and (3) have some evaluation evidence on their positive impact. In the June 3, 1996, draft standards and in this article, "program" will often be used to refer to all these R&D-based models or resources. The OERI legislation includes research findings in this list, but instead of having expert panels designate promising or exemplary research studies, syntheses and interpretive papers, or informational videos, it is most likely that research findings or principles will be used indirectly by including them in the criteria established by each panel.

[4] Merit is determined by expert panel reviews using criteria under the four categories in the OERI standards for designating promising and exemplary programs. These categories are: evidence of effectiveness/success, quality, educational significance, and usefulness to others. (See Figure 2.)

[5] The National Diffusion Network (NDN), established in 1974 and last funded in 1995, provided a systematic procedure to identify and disseminate programs that had evidence of effectiveness. The System of Expert Panels builds on the NDN focus on reviewing submissions to judge effectiveness of replicable programs. But it goes beyond NDN in many ways, such as including three additional criteria categories and deliberately using experts, criteria, and evidence appropriate for specific topic areas. NDN was a self-contained, general-purpose review and dissemination system with a review panel. ED funded many of those approved by the panel as developers/demonstrators (to provide implementation assistance to adopters) and state facilitators (to encourage the use of all approved programs).

[6] This governance structure may be part of NEDS, or it may be created specifically for D+E providers with ties to the expert panels.

[7] Constituency groups as represented by associations often recognize excellence by giving awards to individuals in their field, or they help establish standards and credentialing systems that focus on judging the quality of individuals or organizations in meeting standards of the profession. They have established activities to recognize promising and exemplary programs, but often find it difficult to sustain this type of review effort without external support or clear connections to continued evaluation and dissemination efforts. The U.S. Department of Education has had similar experiences with a wide range of recognition and review activities.

[8] *Consumer Reports* from the Consumers Union could be one model. Each report would contain an overview describing the programs and key issues related to their merit, a comparison chart, and individual program summaries.

[9] Individuals interested in participating in the "Findbest listserv" should send an e-mail message to tara_ariola@ed.gov with name, postal address, telephone and fax numbers, and preferred e-mail address. Individuals interested in participating in "gndrpan" should e-mail sue_klein@ed.gov with the same information, plus information on whether they have expertise in the Gender Equity Expert Panel Subpanel topic areas of core gender equity, gender equity and disability, mathematics, science and technology, prevention of violence and sexual harassment, teacher education and professional development, and vocational education and school to work.

[10] In a more recent article by Robert Slavin (1997b), *Reading by Nine: A Comprehensive Strategy*, he agrees that this U.S. Department of Education System "would be a logical part of a process of identifying programs ready for third-party evaluation" (p. 12).

References

Backer, T. (1995, February). *Readiness for change, educational innovations, and educational reform* (Human Interaction Research Institute). Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Barkdoll, G. (1995). *Standards and their use in the Food and Drug Administration (FDA)* (January 19, 1996). Washington, DC: OERI.

Birman, B., & O'Malley, A. (1996a). *Expert Panel in Mathematics and Science Education: Summary of first panel meeting* (July 25, 1996, draft, Pelavin Research Institute). Washington, DC: OERI.

Birman, B., & O'Malley, A. (1996b). *Expert Panel in Mathematics and Science Education: Summary of the second panel meeting* (November 15–16, 1996 draft, Pelavin Research Institute). Washington, DC: OERI.

Birman, B., & O'Malley, A. (1997). *Expert Panel in Mathematics and Science Education: Summary of the third panel meeting* (January 27–28, 1997, draft, Pelavin Research Institute). Washington, DC: OERI.

Bodilly, S. (1996). *Lessons from New American Schools development corporation's demonstration phase*. Washington, DC: RAND.

Bodilly, S., Purnell, S., Ramsey, K., & Smith, C. (1995). *Designing New American Schools—baseline observations on nine design teams*. Washington, DC: RAND.

Bogart, K. (1996). *Notes on the federal dissemination experts meeting: Advising the Department of Education's Office of Educational Research and Improvement on its design of a systematic consumer-oriented evaluation system to designate promising and exemplary educational products, programs and practices* (October 12, 1995). Washington, DC: Academy for Educational Development.

Campbell, P. B. (1994). *Whose knowledge is it?: Involving teachers in the generating and using of information on educational innovations* (July 10, 1994). Washington, DC: OERI.

Campbell, P. B., & Bachmann, K. A. (1996, November). *First year formative evaluation: The Pilot Gender Equity Expert Panel* (Campbell-Kibler Associates, Inc., Groton, MA). Washington, DC: OERI.

Canine, D. (1995). *Using what we have to get the schools we need: A productivity focus for American education*. New York: Consortium on Productivity in the Schools, Institute on Education and the Economy, Teachers College, Columbia University.

Datta, L. (1994). *A matter of consensus. Datta analysis* (August, 17, 1994). Washington, DC: OERI.

Datta, L., & Scriven, M. (1997). *Incorporating research findings and practices into expert panel work: A dialog between Michael Scriven (MS) and Lois-ellin Datta (LD)* (January 31, 1997). Washington, DC: OERI.

Hanson, K. (1996, April). *Concept paper: Gender Equity Expert Panel* (Rev. October 1996). Newton, MA: Education Development Center.

Hollifield, J., Stringfield, S., & Herman, R. (1996). *Evaluation and standards for schoolwide programs and programs conducted in multiple sites*. Washington, DC: OERI.

Klein, S. S. (Ed.). (1993). Special feature. Sharing the best: Finding better ways for the federal government to use evaluation to guide the dissemination of promising and exemplary education solutions. *Evaluation and Program Planning, 16*(3), 209–278.

Klein, S. S. (1995). When school meets science: The role of the U.S. Department of Education in helping educators use the results of research and evaluation. In T. Tyden (Ed.), *When school meets science* (pp. 160–175) Stockholm, Sweden: Stockholm Institute of Education Press.

Klein, S. S. (1997, January 15). *The evolving "Findbest" system.* (Working Document). Washington, DC: OERI.

Klein, S. S., and Gwaltney, M. K. (1991). Charting the education dissemination system: Where we are and where we go from here. *Knowledge: Creation, Diffusion, Utilization, 12*(3), 241–265.

Komoski, P. K. (1989). Problems with consumer information in education. *Knowledge: Creation, Diffusion, Utilization, 11*(1), 51–57.

Lacampagne, C. (1996, June). *Concept paper: Expert Panel in Mathematics and Science Education*. Washington, DC: OERI.

LaFollette, M. (Ed.) (1992). New models for disseminating education R&D. *Knowledge: Creation, Diffusion, Utilization, 13*(3). Newbury Park, CA: Sage Periodicals Press.

Maxwell, J. (1996, September). *Guidelines for submitting programs to the Gender Equity Expert Panel for review.* Newton: MA: Education Development Center.

Muscara, C. (1996). *A discussion of some U.S. evaluation efforts for programs and resources in mathematics and science* (Computer Technology Services, Inc., October 1996). Washington, DC: OERI.

Schmieder, A. (1996, April). *OERI leadership efforts in designing approaches to identity and share promising and exemplary products, programs and practices.* Notes from an OERI-sponsored meeting before the AERA annual meeting in New York.

Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation, 58*.

Scriven, M. (1995). The logic of evaluation and evaluation practice. In D. M. Fournier (Ed.), *Reasoning in evaluation: Inferential links and leaps. New directions for program evaluation* (Vol. 68, pp. 49–70). San Francisco: Jossey-Ball Publishers.

Seidman, P. (1997). Expert panels for promising and exemplary innovations: A fine idea from the feds. *NCRVE Centerwork, 1*(1), 5–7.

Slavin, R. E. (1997a). Design competitions: A proposal for a new federal role in educational research and development. *Educational Researcher, 26*(1), 22–28.

Slavin, R. E. (1997b). *Reading by nine: A comprehensive strategy*. Manuscript in preparation.

Standards for Conduct and Evaluation of Activities Carried out by the Office of Educational Research and Improvement (OERI)—Designation of Exemplary and Promising Programs: Proposed Rules. 34 C.F.R. Part 701 (June 3, 1996, *Federal Register*).

Turnbull, B. (1996, March). *Technical assistance and the creation of educational knowledge* (Policy Studies Associates). Washington, DC: OERI.

U.S. Department of Education. (1994, December). *Strategic plan for the U.S. Department of Education* (Working Document). Washington, DC: Author.

Statutes

Title IX of Goals 2000: Educate America Act. Public Law No. 103-227, Section 912f (March 21, 1994), 103rd Congress.

PROFESSIONAL EXAMINATION SERVICE

Annual Award for Scientific Contributions to Credentialing

Professional Examination Service (PES) is pleased to announce a call for nominations for the 1997 Award for Scientific Contributions to Credentialing.

Accompanied by a prize of $1,500, the award recognizes scientific or technical research that has contributed to improved measurement practices related to credentialing and to the enhancement of the public-protection function of credentialing programs. The award will be given for contributions including, but not limited to, innovative ways of solving theoretical or practical measurement problems encountered in credentialing programs, creative testing procedures or products, and contributions to the methodologies used in credentialing applications.

Nominations for the award should consist of a detailed curriculum vitae and a 4- to 5-page statement describing the nominee's contributions to credentialing practice. A panel of three non-PES reviewers will judge the nominations on the basis of the quality, significance, and breadth of the individual's contributions.

Nomination submissions and inquiries should be directed to: Karen Cullen, Public Service Activities Manager, Professional Examination Service, 475 Riverside Drive, New York, NY 10115-0089, (212) 870-3494, fax: (212) 870-3333, e-mail: mission@proexam.org. **Nomination deadline is October 3, 1997.**

UPDATE

Calls for Papers

The **Association for the Study of Play** will meet in St. Petersburg, FL, February 18–22, 1998, in conjunction with the Society for Cross-Cultural Research. Proposal deadline: October 15, 1997. Contact: Dorothy J. Sluss, 70548 Human Development and Learning, East Tennessee State University, Johnson City, TN 37614; phone 423/439-6342; fax 423/439-4439, e-mail sluss@access.etsu-tn.edu.

Calls for Reviewers

The **National Head Start Bureau** is expanding its pool of peer reviewers and consultants to assist with the review of current and future activities, policies, and research. Contact Dakota Technologies Corporation, Head Start Reviewers and Consultants, A-E7, Fair Oaks Commerce Center, 11320 Random Hills Road, Suite 105, Fairfax, VA 22030, fax 703/218-2483, e-mail dakotacorp@aol.com.

Calls for Essays

The **Institute for Scientific Information** is soliciting 500-word essays from researchers who have used Current Contents, the current awareness database, about how it has been valuable to their research. The winner of the grant will receive $4,000. Deadline: September 30, 1997. No entry form is required. For information, contact: Evette Fearon-Bennett, Associate Marketing Manager, Institute for Scientific Information, 3501 Market Street, Philadelphia, PA 19104-3389, fax 215/387-1125, e-mail efearonb@isinet.com or visit http://www.isinet.com.

Calls for Applications

The **Law School Admission Council Research Grant Program** funds empirical research on legal education in the U.S. and in the common law schools of Canada. There are two reviewing cycles for these grants each year. Deadlines are September 1, 1997, and February 1, 1998. Contact Kath-

continued on p. 45

Rejoinder:

Design Competitions and Expert Panels: Similar Objectives, Very Different Paths

ROBERT E. SLAVIN

Sue Klein's responses to my article on design competitions (Slavin, 1997) provide an excellent summary of the U.S. Department of Education's plans to replace the now-defunct National Diffusion Network (NDN) with a mechanism to identify, recognize, and possibly fund dissemination of proven and promising educational products. While the details of the System of Expert Panels and of other attempts within the federal government to disseminate effective programs have not yet been established, Klein's article presents an outline of Department of Education thinking in this area.

Clearly, American education needs both a system to recognize and disseminate effective programs, like the proposed System of Expert Panels, and a system to produce new methods and to rigorously evaluate new as well as existing programs, like design competitions. These two proposals do resemble each other in overall objectives and in some particulars, as Klein notes. Yet they are enormously different in scale. The System of Expert Panels would be, as I understand it, an extremely low-budget operation. It would collect and summarize opinions from users of educational products, and occasionally commission consensus panels to evaluate the evidence supporting various instructional approaches. Funding might be made available for exemplary programs, although the details of this are currently unclear.

If the federal government is determined to spend very little on creation, evaluation, and dissemination of educational programs, the System of Expert Panels may be the best system we can expect. This plan also makes sense if we assume that there are dozens or hundreds of truly effective, rigorously evaluated educational methods out in the schools that only need to be brought to the attention of educators everywhere.

The assumptions behind design competitions are quite different. First, I do not believe that there are more than a handful of replicable programs that meet the most minimal standards of evaluation. In a recent review with my colleague Olatokunbo Fashola (Fashola & Slavin, in press), we combed the educational literature trying to find elementary and middle school instructional programs that had been evaluated in comparison to matched control groups on measures of academic achievement and had been replicated beyond their pilot sites. Across all subjects and grade levels, we found only 17 programs that met these minimal standards. To be able to list more than one or two programs in each category, we had to reduce our standards to include "promising" as well as proven programs. If we had included the common-sense requirements of replicated evaluations or evaluations by third parties, our list would have fallen to at most three to four programs. (In contrast, standards for identification as "exemplary" by the National Diffusion Network were so low that by the time its funding ended, more than 500 programs had qualified.) It took us a year to do this review, by the way, and thousands of person-hours. In contrast, Klein's proposal is to have unpaid volunteers review the evidence submitted to them. This process is likely to reflect the current zeitgeist about what effective practice should be, rather than a careful reading of the actual evidence; just a few years ago—and possibly today, depending on who was chosen for an Expert Panel—any beginning reading program emphasizing phonics could have been rejected as "promising" or "exemplary" (regardless of the strength of its evaluation) on the basis of the "quality" category, which is substantially a judgment of the current political correctness of a program.

The idea behind design competitions is to quickly and dramatically increase the number of proven, replicable programs, providing substantial funds for the creation of new, replicable programs, third-party evaluations of these as well as existing promising models, and dissemination of programs that meet rigorous evaluation standards. The ultimate goal of this enterprise is to create a situation in which school staffs can make informed choices from among programs highly likely, if well implemented, to significantly accelerate student achievement. Until we have such a set of proven models, we will continue to jump from fad to fad. We can never have meaningful progress in educational programs until we can have the same level of confidence in them that physicians can have in procedures or medications passed by the Food and Drug Administration.

The one danger in the System of Expert Panels is that it is so inexpensive and minimalist a strategy that it could inhibit more serious attempts to build a system based on dissemination of rigorously evaluated methods. This was essentially the fate of NDN. NDN never had anything near the amount of funding needed to accomplish its very ambitious goals. With a total budget at the end of about $13 million, it provided a single state facilitator for each state and small grants to a few dozen "developer/disseminators" to serve a nation of 100,000 schools in 15,000 districts. Program Effectiveness Panels (and before that, Joint Dissemination Review Panels) were composed of already overworked individuals from across the Department of Education who had to make decisions based on whatever evidence was submitted to them. If a program was used in a hundred schools, a developer could nevertheless submit evidence showing large NCE gains in a single year in just a few of these schools and qualify as an "exemplary" model. This process was much better than nothing in that it at least disseminated information on programs in which there was interest in replication. Because of NDN's existence, however, policymakers felt that they had taken care of the need to disseminate effective programs. Yet, because of inadequate funding, NDN's impact was perceived to be small, despite studies showing its extraordinary cost-effectiveness in terms of numbers of schools served per dollar spent. Because of NDN's low standards of evidence, its programs lost much of their credibility among educators. One hopes that the System of Expert Panels can avoid some of these pitfalls, but the history of NDN provides a cautionary tale about how a low-budget recognition and dissemination program can inhibit development of an adequate program and can ultimately be self-defeating.

Done well, a system of design competitions could cost more than $100 million per year, about eight times NDN's funding, but less than 2% of Title I funding and an infinitesimal proportion of all educational funding. Klein's proposed System of Expert Panels would be an important part of a comprehensive strategy of disseminating educational programs, and it may be a necessary stopgap until we are willing to commit the resources necessary to fund something like design competitions. But let's be clear about the limitations of this approach. It is no substitute for third-party evaluations of well-designed, replicable programs able to help all children meet demanding standards of achievement.

Notes

This article was written with funding from the Office of Educational Research and Improvement, U.S. Department of Education (Grant No. R-117D-40005). However, any opinions expressed are those of the author and do not necessarily represent the positions or policies of the U.S. Department of Education.

References

Fashola, O. S., & Slavin, R. E. (in press). Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk.*

Slavin, R. E. (1997). Design competitions: A proposal for a new federal role in educational research and evaluation. *Educational Researcher, 26*(1), 22–28.

FALL and WINTER RESEARCH INSTITUTES

"Using a Journal as a Qualitative Research Technique: Problems, Perspectives, and Possibilities"

Director: Valerie J. Janesick, FL International University; November 22–23, 1997; Washington, DC; $165 AERA Members, $195 Nonmembers

The notion of using a comprehensive reflective journal to address the researcher's self is critical in qualitative work because the researcher is the research instrument. Journal writing is only now being used and talked about as a serious component in qualitative research projects. In this institute, journal writing will be viewed as a major source of data. Study areas include problems of representation, co-construction of meaning with participants who keep a journal, interpretation of each other's data, gender issues, class issues, racial issues, and therefore, the politics of research. Often writers/researchers are positioned outside the very people and situations they are writing about. Journal writing personalizes representation in a way that forces the researcher to confront issues of how a public text legitimates itself and makes claims for its own authority. Members will practice journal writing and deconstruct the complexities of journal writing as a research technique, including moving from the field to the text to the final public research report; problems of interpretation, meaning, and representation; and issues of race, class, and gender. Other applications and uses of journal writing will be discussed in terms of usefulness to and for educators. Participants should first read Ira Progoff's (1992) **At a Journal Workshop,** Los Angeles, CA: P. Tarcher, Inc.

"Arts-Based Approaches to Educational Research"

Directors: Tom Barone, Arizona State University; Elliot Eisner, Stanford University; Instructors: Lynn Butler-Kisber, McGill University; Jean Clandinin, University of Alberta; February 5–7, 1998 (2.5 days); Tempe, AZ; $195 AERA Members, $230 Nonmembers

During the past two decades, qualitative approaches to the study of educational practice have become increasingly important. Most of these approaches have resided within a social science framework. Ethnography, for example, has been a primary source for educational researchers seeking qualitative methods for revealing the details of educational life. This institute is designed to explore arts-based approaches to the study of educational practice, especially those related to educational connoisseurship and educational criticism. Taking its lead from what critics do in studying works of art, literature, film, music, and theater, educational criticism is designed to exploit the capacities of perception and the artistic renderings of language to describe, interpret, appraise, and thematize the objects and events studied. The institute will be aimed at enabling educational researchers to understand the epistemology underlying arts-based approaches in general and educational criticism in particular. The institute staff will examine issues such as validity, generalization, and artistically rooted forms of understanding. In addition, the institute will afford participants an opportunity to utilize arts-based methods for critiquing real-time videotapes of teachers working in elementary and secondary school classrooms.

To receive an application for the AERA Fall or Winter Institutes, contact: **Fall or Winter Institutes, 1230 17th Street, NW, Washington, DC 20036-3078; phone 202/223-9485; fax 202/775-1824. Participation is limited. Apply early!**

Z

April 6, 1998 Draft

Agenda: The Evolving System of Expert Panels

Sponsored by the Office of Educational Research and Improvement, OERI

Mon. April 13, 8:30-11:30 AM, Rancho das Palmas, Level 4 South Tower Marriott, San Diego

8:30-9:30 ***Current Status and Plans for the System of Expert Panels***

8:30-8:40 Welcome to the Session: Peirce Hammond, Director of the Office of Reform Assistance and Dissemination, OERI

8:40-8:50 ED Precedents to this System of Expert Panels:
John Evans, Developer of the Joint Dissemination Review Panel
Chris Dwyer, RMC ,Contractor for the Program Effectiveness Panel & NDN

8:50-9:00 The Current Status of the Expert Panels: Sharon Bobbitt, Director, Knowledge Applications Division, Office of Reform Assistance and Dissemination

9:00-9:10 How The System of Expert Panels Differs from Past Efforts and How it Can Address Current Challenges: Sue Klein, Acting Team Leader, System of Expert Panels, Knowledge Applications Division

9:10-9:20 Overview of the Mathematics and Science Expert Panel:
Maria Lopez-Freeman, Panel Member
Susan Klaiber, RMC Panel support contractor

9:20-9:30 Overview of the Gender Equity Expert Panel: Panel Co-Chairs
Harilyn Rousso, Disabilities Unlimited
Mary Wiberg, Iowa Dept of Education

9:30-10:40 ***Discussion Groups:***

Procedures and Room Locations: Susan Klaiber

Group Leaders give 3 minutes on their topics and what they hope to learn

Group 1: Advice on Creating a Long Lasting and Effective System of Expert Panels: Peirce Hammond, OERI, ORAD

Group 2: Advice on Distinguishing Between Promising and Exemplary Interventions: Gary Borich, Univ. of Texas and a member of the System of Expert Panels new Impact Review Panel and Lynn Fox, American Univ. Chair of the Gender Equity Expert Panel's Subpanel on Mathematics, Science and Technology

Group 3: Advice on Guiding Future Submitters to the Panels: Lois-ellin Datta, Data Analysis, and Patricia Campbell, Campbell-Kibler, Assoc.

10:40-11:30 ***Summary Reports from Each Group and General Discussion***

The Evolving System of Expert Panels
A Presession on April 13, 1998 at the
American Educational Research Association Conference
San Diego Marriott Hotel and Marina
Sponsored by the Office of Educational Research and Improvement, OERI
U.S. Department of Education

The session was opened by Susan Klein, Acting Team Leader of the Expert Panel System who introduced Peirce Hammond, Director of the Office of Reform Assistance and Dissemination, OERI. Hammond welcomed the group of about 30 people and noted that the Expert Panel System is at the heart of what OERI is all about. The Office of Educational Research and Improvement is about knowledge generation and knowledge use and links research with use, action and practice. The Expert Panel System is the latest attempt at achieving this goal.

Two Department of Education precedents to the Expert Panel System were then briefly described. John Evans, a current member of the Impact Review Panel and a developer of the Joint Dissemination Review Panel noted that the current efforts are not original. He went on to explain that, 20 years ago, the Office of Education was promoting programs for which there was no evidence of effectiveness. This led to the origination of the Dissemination Review Panel (DRP), which soon became the Joint Dissemination Review Panel (JDRP) when it joined efforts with the National Institute of Education. The DRP, and then the JDRP, was an effort to put in place a quality control mechanism and federal dissemination was not allowed unless a program/project had been approved by the Panel. Administered by a half time person, Evans pointed out that the system was a model of administrative simplicity and economy.

Chris Dwyer, Senior Vice President of RMC Research and a long-time technical assistance provider to the JDRP and its successor, then described the transition, in 1986, to the Program Effectiveness Panel. Changes to the system included panel membership moving from totally federal employees to a requirement that at least two-thirds of the members be non-federal personnel and that programs to be considered by the Panel could be funded through a variety of sources -- not limited to federal funding alone. The Panel continued to review only the evaluation design and results of a program. Congress did not allow the Panel to review the content of a program. While the early years of the Panel were characterized by review of classroom-developed programs/projects, by 1995 most programs being reviewed by the Panel had been developed by universities, Labs, foundation supported enterprises, and other non-profit entities. Dwyer noted that, over the years, there has been a shift in beliefs about how people learn. In the early days of the JDRP, the PEP, and the NDN, replication of proven practices was seen as the logical best way to change what was happening in schools and classrooms. By the late 1990s, education has shifted to a more constructivist approach. Dwyer concluded by saying that beliefs about how people change and how people learn have an influence on the development of the Expert Panel System.

Sue Klein provided the group with an update on the current status of the Expert Panel System and also remarked about how the current system differs from past efforts and how it can address current challenges. A handout titled *System of Expert Panels* was provided to participants which describes the current status of each of five Expert Panels including Mathematics and Science; Gender Equity; Safe, Disciplined, and Drug-Free Schools; Educational Technology; and Early Reading. The Mathematics and Science Panel and the Gender Equity Panel are currently reviewing programs. The other three panels are in varying stages of development. Klein also noted that there are other parts of the Department considering the establishment of panels including racial and ethnic equity, OBEMLA, and Adult and Vocational Education. It was also pointed out that there is a desire for some basic consistency across all Expert Panels but that there is a need to allow some differences by panel. Areas of consistency include 1) the four legislatively required broad criteria of quality, replicability/usefulness, evidence of effectiveness, and educational significance and 2) the Impact Review Panel which reviews the evidence of effectiveness across all panels.

A member of the Mathematics and Science Education Expert Panel, Maria Lopez-Freeman, gave an overview of the Panel's developmental activities noting that time was spent, over the last one and one half years, determining the criteria and indicators, discussing the difference between promising and exemplary, and generally establishing panel procedures. Input from the California textbook adoption experiences and the development of the AAAS Benchmarks made important contributions to the current procedures being used. Susan Klaiber, RMC Research's director of the technical assistance contract for the System provided an update on the review of mathematics programs that is currently underway. A total of 61 programs have been reviewed by field-based reviewers. Those programs meeting a high quality standard are being referred to the Impact Review Panel for review of the evidence of effectiveness. The full Math/Science Panel will meet at the end of May to make recommendations of programs to be designated promising and exemplary.

Harilyn Ruosso, co-chair of the Gender Equity Panel provided background on the development of the Gender Equity Panel. Over the past two years the Panel has established its procedures and is currently reviewing a group of submissions. The Gender Equity Panel chose to organize itself around six subpanels including Prevention of Violence and Sexual Harassment; Gender Equity and Disability; Mathematics, Science, and Technology; Teacher Education and Professional Development; Vocational Technical Education/School to Work; and Core Gender Equity. Each subpanel made contact through its own network to solicit submissions. A total of 20 submissions was received and are in varying stages of review by the subpanels and the full panel. The Gender Equity Panel had established their panel procedures before knowing about the Impact Review Panel so the various subpanels are reviewing the evidence of effectiveness for each program. All programs determined to be exemplary will be forwarded to the Impact Review Panel for evaluation and comment with final recommendations of programs to be designated promising and exemplary to be made by the full Gender Equity Panel.

After hearing the background information on the development of the System, opportunity for discussion in smaller groups was provided. The three small group discussions included: Advice on Creating a Long Lasting and Effective System of Expert Panels; Advice on Distinguishing Between Promising and Exemplary Interventions; and Advice on Guiding Future Submitters to the Panels. Notes from each small group follow.

Notes from Discussion Group 1:
"Advice on Creating a Long Lasting and Effective System of Expert Panels"
Peirce Hammond, Director, ORAD, OERI

Attendees: 14

Discussion

Peirce Hammond: Introduction

The topic is one of discussing the "architecture" of the system and numerous questions can be considered such as:

- What is "longlasting?"
- How activist should the panel be?
- What policy should the Department have on promising vs. exemplary; How is that designation distinguished from other people's use of the terms?
- How can the system sustain itself?

Are there other questions to consider?

Katherine Hanson, ERC, EDC:

"What's in it for us ?" (meaning those who might/could submit) There needs to be a connection between equity and the Panel System

Frank Murray, University of Delaware and Impact Review Panel member:

Concern about every AERA SIG wanting its own panel; how does the system deal with an overabundance of panels? Need to deal with legitimacy issues and the resources available; it is a "grain size" issue

Gene Hall, University of Northern Colorado:

How are we going to learn from the process?

Peirce Hammond question to John Evans:

What should not be repeated and what features should be maintained from earlier efforts?

John Evans, San Joaquin Delta College and Impact Review Panel member:

Concerned about the potential for disabling diffuseness; the System needs credibility with Congress; need to make a decision about the number of panels; right now there is no sense of a consistent core; recommend that PH be the "czar of the Panel;" evidence of effectiveness is the *sine qua non* of the Panel System. With substantive panels, panelists are chosen as supporters for programs in their area; there is a potential problem of subordinating the issue of effectiveness.

Margaret Camarena, McREL:

What is the mechanism to disseminate? Use the Labs? Other mechanisms?

Gene Hall:

Concern that "Having lost sight of the objectives, they redoubled their efforts." There is a need for balance between how much effort and rigor goes into this part of the process.

Frank Murray:

What is the "purpose?" Is the plan to cast a wide net? *Consumer Reports* goes out and gets what they want to evaluate. Is the public being misled? Can the government help us make decisions?

Peirce Hammond:

Having identified some more of the issues, let's move to discussing some of the questions. How do we create a balance between having a place for all programs to be reviewed and limiting the number of panels?

Gene Hall:

It has been said that everything done so far is a pilot. Label it as such and move to a generic panel as a pilot.

John Evans:

There is no need for more pilots. There need to be some decisions made. How many panels will there be and what will they be? What will the role of the individual panels be relative to substance vs. effectiveness.

Katherine Hanson:

There is a need for activism. How can there be a broader understanding of what works for whom? What are the criteria for good programs? Maybe there is a need for cross-panel discussions regarding certain issues such as diversity, gender equity, etc.

Peirce Hammond:

How many panels?

Sue Klein:

We need a list of criteria to establish a panel, e.g., Is there a strong constituency working in R&D in an area? All Institutes should be looking for what works in their area; Is there a demand from the field? What if an area is very controversial, e.g., Social Studies; perhaps consider other areas where there are funds and/or needs, e.g., 21st Century Schools (after school programs). In summary, panels may be established based on a list of specified criteria and there may be some panels formed that are opportunistic.

Sandy Berger, McCREL:

Perhaps there should be a panel for each Institute.

Sue Klein:

That had been considered but the Institutes are so diverse that it did not seem to make sense.

Peirce Hammond:

There might be a natural formation of panels around the basic content areas, e.g., reading, math, etc. and around programmatically driven areas, e.g., technology, 21st Century Schools, etc.

Frank Murray:

From the field's perspective the questions are: What is the news? and What is truly innovative?

Peirce Hammond:

We have three panel types within the five panels already established:

Content: Mathematics and Science, Early Reading
Programmatic: Safe, Disciplined, and Drug-Free Schools, Technology
Crosscutting: Gender Equity

How do we keep it from becoming "too many" panels?

Suggestion: Perhaps there could be 1) core panels, probably subject matter based and 2) temporally bound panels.

Gene Hall:

What about considering defining panels in terms of student outcomes? e.g., changes in school culture or gains in a particular area.

Notes from Discussion Group 2:
"Advice on Distinguishing Between Promising and Exemplary Interventions"
Gary Borich, University of Texas and Lynn Fox, American University

Attendees:
Gary Borich, Angela Chan, Tony Eichelberger, Lynn Fox, Maria Lopez-Freeman, Tom Owens, Sylvia Rosenfield, Joan Ruskus, Peggy Simon, Bernadine Stake, Flo Stevens, Veronica Thomas

Discussion

GB: Introduction

Qualitative vs. Quantitative Concerns and Issues:

- Tension exists between agencies wanting quantitative data, but this tension doesn't exist in the field. There's lots of qualitative data (e.g., case studies, etc.); and there's some quantification with this, but very little significance testing.
- Needs to be balance between qualitative and quantitative. One pole isn't better then the other.
- Vague distinctions between programs, practices, policies, products.
- Single standard would not work with all of above. Need different standards.

Goal of this Session:

- Identify what are legitimate concerns and clarification of these concerns.

TE:

- Use of a program in the field may be less than the utilization of a whole program and may be more a selection of some parts of a program to apply to the setting.
- Evidence of effectiveness is an issue if you're just looking at parts – evidence the program cites may be applied to the whole program and not necessarily a particular component of the program

GB:

- Programs lose their integrity from the original data collected when components are selected.
- Opportunity to collect data is difficult – programs, products, policies. Policies and practices often linked to systemic systems (i.e., they work in the system).

FS:

- The context changes across sites. What's important is the context (e.g., conditions) under which this program, practice, etc., works.
- For example, "Success for All": There is a system of implementation set-up, but it changes depending on the context in which it will be implemented. Need to know how and with whom (e.g., populations) this has worked.

BS:

It's okay to change the program depending on the context.

FS:

If you say something's good, need to specify under what conditions.

SR:

Need to know the critical elements that are needed to actually implement the program so you know you are implementing what is necessary to implement (e.g., system, setting, population).

GB:

- Other ways of knowing – What is the program? What is the independent variable?
- Programs only provide cursory description of the program.
- What is actually being implemented and evaluated?
- Programs are multi-faceted.
- Need standards for more thorough description/articulation of different components; and under what circumstances the components are implemented.

JR:

Quantitative vs. Qualitative: Qualitative – need to identify what the independent variable is and tease out the various factors. This is a good way the qualitative research can be used.

TO:
If there's no evaluation impact or evidence, you just have judgment that program has quality.

JR:
Quality of program needs to be independent of judgment of evidence of effectiveness.

TO:

- Consumer-Oriented Evaluation: Different approach/goal than past validation systems (e.g., PEP, JDRP).
- More important, these approaches have two different purposes. The consumer protection approach looks at descriptions of weaknesses as well as strengths of a program. The advantage is not just quality of the product but establishment of a dialogue among a variety of users.
- New Purpose: Continued improvement of "math" field through dialogue with users and developers. "Community of learners" applied to this process.

GB:
Is this the role of the promising designation?

TO:
Could be the role for both exemplary and promising designations.

GB:
Promising – do not meet all exemplary standards, but being in this category would put you in competition for resources needed to achieve exemplary status.

FS:

- The exemplary programs will be disseminated and not just those producing the product could get additional resources.
- Pluses and minuses (even if exemplary – who it is serving and who is providing services) need to be known so others know if it will work in their context and what supports are necessary to use this program.

SR:

- Hard for schools to know how to make decisions about what programs to implement (e.g., referring to the consumer model).
- How can rubrics or frameworks to guide schools' decisions to choose programs that will work for them be set-up?

TO:
Find out experiences of the program in different settings.

SR:
What do schools need to know about a program to know if it is appropriate/will work for them (e.g., what parts will work).

TO:

- What's value-added from your experience that will improve what I already have in place?
- Do you need to throw out your whole/old program?

JR:

- What is the consumer-oriented approach?
- Who is the consumer and what is to be disseminated?
- What the government defines is not what we are talking about.
- Pendulum swinging back to quantitative/control group and away from the qualitative approach.

TE:

- Pressure has always been to relieve need for quantitative approach and allow for more qualitative data responsive to what educational programs can provide.
- Can do description under what conditions program works and where the problems are, etc. But I haven't seen "good" qualitative work that also goes beyond personal report.

JR:

This is coming as work goes on.

TO:

Is there intent that Congress will approve dissemination funds for these programs?

GB:

- There is no money from government, but the designation will help programs obtain resources from other sources.
- Which types of programs will respond to this "carrot"?

FS:

The "carrot" will not be useful to classrooms necessarily.

TO:

What is the purpose of the Expert Panel validation system? Ought to look at fresh purposes responsive to education in a broader sense.

LF:

Some programs look to the external validation to help them continue their work and access evaluation.

FS:

Problem of which comes first – "the chicken or the egg"? If programs don't have evaluation, can they gain validation needed to continue and gain other resources?

SR:

Many universities have not focused on qualitative evaluation and remain focused on quantitative methods of program evaluation.

JR:

But some universities are broadening their focus.

GB:

- Who are we training (i.e., what audiences) to understand qualitative as well as quantitative methodologies. We need more training of administrators and developers of programs.
- Federal government is a little more in the quantitative camp, but needs to understand the diversity of contexts out there.
- Limiting exemplary status to more hard quantitative data may eliminate programs that have good qualitative evidence.
- Both qualitative and quantitative methodologies are needed.

FS:

- But the thought out there (e.g., federal government, NSF, school districts, public, parents) is the need to show hard quantitative evidence of student improvement in knowledge base.
- We can push for qualitative approaches but academic achievement is the bottom line.

TE:

What would be helpful is having different types of claims/exemplars available (e.g., similar to PEP).

JR:

People honed in on quantitative methods need to pay attention to violations of the assumptions of quantitative methods.

Summary

GB:

- Concern about qualitative data being difficult to interpret unless presented in a quantitative context.
  - ★ Need balance between the quantitative and qualitative approaches.

- Practices, programs, products – need different approaches to evaluate these very different types of programs (e.g., policies can only be evaluated in the context of their systemic structure in which this resource fits).
  - ★ Need articulation of the context.

- How can qualitative approach be used? Even if a program is exemplary, it still has flaws.
  - ★ Can use qualitative approach to describe the "real-world" events (e.g., document the things that go right or wrong and why and what parts work in particular settings).

- Replicability – the extent to which the program is used and in what sites.
  - ★ Valuable source of information is meta-analysis of how this program is used in many different sites by different people.
- When something is exemplary and passes by quantitative standards, often this does not look at

violations of the assumptions of quantitative methods.

- ★ To what extent does the quantitative evaluation meet the assumptions of quantitative methods?

- What is the role of qualitative analysis?
  - ★ It can carefully define the program itself. There are many "bundles" of independent variables (i.e., components of the program). Qualitative analysis can help understand what these program components are.
  - ★ It can define the implementation in the "real-world". Can describe what were the levels of implementation of the different components.
  - ★ It can confirm quantitative results (e.g., vignettes can extend and support the results of the quantitative analysis).

- The hope is that the pendulum is in the middle of the qualitative/quantitative debate for the Expert Panel.

LF:

★ What is the consumer-oriented approach?

★ People trained to do analysis differ from those "in the trenches".

★ Evolutionary nature of the Expert Panel approach – a new purpose.

★ Can this inform audiences of ways to improve the field?

Notes from Discussion Group 3:

"Advice on Guiding Future Submitters to the Panels"

Patricia Campbell, Campbell-Kibler Associates and Lois-ellin Datta, Datta Analysis

---

This discussion about the format and contents of written guidance for Expert Panel applicants was co-led by Lois-ellin Datta and Patricia Campbell and who began the session by identifying some key questions to be discussed by the group: What is needed to consider making a submission, including the resources required from a funding perspective? What level of detail is appropriate? How do you handle information about what has not worked along the way in project development? The co-leaders also added some concerns for the group to ponder: the blurring of the distinction between evaluation and research; how to convince programs that they may have more data than they realize and communicate how to pull information together; the difficulty of finding measures that are sensitive to change, i.e., measures that will reflect change if it occurs even in a short-term evaluation.

Participants arrived at the following recommendations:

--provide lots of examples of claims and evidence, and include examples that illustrate how to weave data into an argument;
--include successful and unsuccessful examples and discuss how to improve upon weak approaches;
--address the issue of program submittals that combine pieces of other existing programs, including a discussion of implications for use of evidence from others' programs;
--encourage applicants to provide context and history;
--be sure that panelists provide feedback to applicants who are not successful about how they might improve;
--provide copies of successful submittals along with technical assistance to prospective applicants;
--include examples of programs at all different levels of the system, i.e., classroom, school, district, state, federal;
--illustrate evidence that is qualitative as well as quantitative, especially in looking at ways to provide evidence for policy use; a suggestion was made that qualitative designs might be used to present the levels of program implementation with qualitative confirming results; and
--funding sources need to provide information about planning for evaluation as early as possible in the life of projects.

Everyone agreed that they did not want a guide by way of examples to be limiting of types of applicants or nature of evidence.

During the discussion, several of the participants were concerned about different fields being at a disadvantage, specifically there were concerns that the gender equity applicants would not know how to pull data together in the way that many math and science applicants would. The other concern related to gender equity was the number of applicants that were likely to be presenting events or repeated "one shot" activities.

There was recognition that it is easiest obtain evaluation evidence at the micro-level, i.e., classroom units or events, but that the goal should be to encourage programs at all levels. Further, the group affirmed that an overall value of the consumer-oriented dissemination process was that it would inform people about evaluation.

Discussants recognized the importance of providing guidance in a way that encouraged worthy programs to develop submittals while at the same time advising those without appropriate evidence not to expend the resources and time to prepare submittals.

AA

The Expert Panel System is a Reasonable Way to Identify Promising and Exemplary Programs— A Response to Stanley Pogrow's article, "What is an Exemplary Program and Why Should Anyone Care? A Reaction to Slavin and Klein"

Susan S. Klein[1]

This response focuses on four issues most relevant to the evolving System of Expert Panels (Klein, 1997). I disagree with the first two of the following points from Pogrow's article and agree with the last two.

1. **Pogrow objects to having a wide variety of expert panels**. Pogrow said **"even if there were lots of great programs, Klein's use of a wide variety of panels to identify such programs is flawed."** His primary concern seemed to be that it would be difficult to find panel members who were "non-ideological" and technically expert. While the panels are still in their early stages, so far the U.S. Department of Education (ED) is having no difficulty in assembling broadly representative panels of highly qualified practitioners and researchers/evaluators who are pleased to serve as national experts in their specialty areas.

   In Pogrow's criticism of the U.S. Department of Education's plans to establish "a wide variety of panels" (perhaps in contrast to one multi-topic panel) he failed to recognize the advantages of multiple long-term expert panels to accomplish our shared goals of development and identification of exemplary programs. Anticipated benefits include:

   Providing a consistent way to involve many experts across the nation in guiding decisions about R&D investments and use: Each expert panel can serve: 1) as a catalyst for the continued involvement of many constituents in the panel's topic area and 2) as a credible source of evaluative information to help consumers decide how to use their local funds to choose among promising or exemplary programs. Initial experiences with the Mathematics and Science and the Gender Expert Panels show that they have involved a wide variety of experts in their fields in and outside federal and state government as reviewers, developers, dissemination partners and general panel advisers/supporters as well as panel members. Each panel provides opportunities for major impact on federal funding decisions and on consumer program selection decisions. For example, the Gender Equity Expert Panel helps ED's Office of Educational Research and Improvement (OERI) implement its gender equity mandate. Research has informed us about gender inequities, but little has been done to identify effective program and policy solutions to advance gender equity. Similarly, the Safe, Disciplined and Drug-Free Schools Expert Panel helps the Department bring together practitioners and individuals with research and evaluation expertise from within and outside the Federal government to identify what works best in given circumstances and contexts.

---

[1] Dr. Klein works in the Office of Educational Research and Improvement in the U.S. Department of Education. This response reflects her private views. It is presented to encourage understanding of OERI efforts to develop a System of Expert Panels, but it is not an official statement by the U.S. Department of Education. In preparing this response, Dr. Klein also saw Dr. Slavin's 8/98 prepublication response, "Yes, Control Groups are Essential in Program Evaluation: A Response to Pogrow". This response includes suggestions from Peirce Hammond and Sharon Bobbitt in OERI; Chris Dwyer at RMC Research; John Evans, Impact Review Panel (IRP) member and developer of the Joint Dissemination Review Panel; and Charles Haughey, Vice-President of the Knowledge Utilization Society. Extensive thanks go to Lois-ellin Datta, IRP member and adviser to the System of Expert Panels for her long term and detailed discussions of the issues and concepts and for her concise editing suggestions.

Evaluation capacity building: Experience with the first round of submissions from the two initial panels shows the sparsity of adequate evaluation evidence to support claims of positive impact. The broad involvement of government and foundation staff, reviewers, submitters, technical assistance providers and others should increase attention to (and support of) this necessary evaluation work in topic areas covered by each panel. It is also hoped that the consumers who select the promising and exemplary programs will become active users of the evaluation information developed by the panels as well as future contributors of evidence based on their own evaluations of the promising or exemplary program they selected.

Building and sharing knowledge of solutions in the topic area: After several review cycles, the panel members should become increasingly able to make informed comparisons of programs with similar purposes and they should be able to identify gaps in the range of available solutions. This type of detailed knowledge of solutions in a topic area is not feasible in a multi-topic review panel or in the typical short-term efforts to compile catalogs of what works in specific topic areas.

Since each topic-focused expert panel is part of the larger System of Expert Panels (the System), there is a need for the panels to work cooperatively and learn from each other so that the whole will be more worthwhile than the sum of its parts. Initial coordination activities include: an Expert Panel web page, <http://www.rmcres.com/expertp/index.html>; the development of a System-wide guidebook for submissions and reviews, an Impact Review Panel[2], and joint meetings of representatives from all panels.

The System of Expert Panels has progressed since my 1997 *Educational Researcher* article. In addition to the first-cycle review work of the Expert Panels on Gender Equity and Mathematics and Science, the Expert Panels on Safe, Disciplined and Drug-Free Schools and on Education Technology have had their initial meetings. Under OERI leadership, the Department is thinking about working with a variety of constituencies to establish new expert panels based on the following considerations:

- Does the panel focus on an area where the Department of Education has legislative responsibilities? (For example, ED has no responsibility for religious education, and other agencies have clear leadership responsibilities for medical education.)

---

[2] An Impact Review Panel, composed of evaluation experts advises the Expert Panels on using rules of evidence to support claims of positive impact to decide what to recommend to the U.S. Secretary of Education as exemplary.

- Is there an external constituency of R&D producers and practitioners interested in learning what works well in their field and who want to play an active role in this partnership? (Are there interested associations and networks of researchers and practitioners in the topic area?)

- Is there a sufficient body of replicable programs or interventions and evaluations in the topic area that will yield products, programs, policies and practices that might be judged promising or exemplary?

- Would ED sponsorship of a panel foster cooperative work within ED and with other Federal agencies with related responsibilities?

- Is it sufficiently different from other existing expert panels in the System?

In summary, there are sound constituency relations and technical reasons for ED to help educators learn about the solutions in topic areas of special concern to them by establishing a coordinated system of multi-year expert panels composed of stakeholder/practitioner and R&D experts. Those involved have been learning from the work of multi-topic review panels, such as the Joint Dissemination Review Panel and its ED successor, the Program Effectiveness Panel as well as from many other Federal and non-Federally sponsored review efforts. Commissioned papers and other documents relating to the System of Expert Panels are available in three volumes of working papers compiled by (Klein, 1997a, 1998). OERI also has a "Findbest" listserv[3] for individuals interested in the System.

2. **Pogrow disapproves of the System of Expert Panels since currently there are relatively few exemplary programs.** In doing so, he fails to see merit in having expert panels identify **promising programs** and in providing incentives and help for developers to obtain evidence to substantiate claims of positive impact to potentially identify more exemplary programs. The System of Expert Panels has been wisely asked by Congress to identify both promising and exemplary programs using an objectively managed process with review criteria that will be meaningful to potential users. Four criteria categories have been established.[4] They are *(a) Evidence of success; (b) Quality of the program; (c) Educational significance; and (d) Replicability*. Each expert panel establishes its own detailed subcriteria for each category. To qualify as promising, the panels must judge the program favorably on all four criteria categories. In the category of evidence of effectiveness/ success, the regulations for the System specify that a promising program must have evidence that it worked in at least one place and an exemplary program must have a positive impact in multiple sites. The "promising" designation is useful as an intermediate designation since Pogrow, Slavin, I and many others agree about the difficulty in identifying exemplary programs, due in great part to

---

[3] If you would like to subscribe to the Findbest listserv, send an e-mail request to <Tara_Ariola@ed.gov>. Include your name, e-mail and work mailing address and telephone number.

[4] These criteria categories or standards are specified in the U. S. Department of Education's Nov. 17, 1997 *Federal Register* regulations for the System of Expert Panels.

inadequate investment in evaluations of program impact. Thus, a "promising" designation is an important way to systematically identify programs worthy of additional investment in their evaluation[5] and refinement.

Pogrow also said, ***"An exemplary program is one that can increase learning to a surprising extent with a great deal of consistency."***(p.22) With regard to exemplary status, I agree with the spirit and underlying concept of the ***"to a surprising extent"*** part of Pogrow's definition and with his focus on consistency which will be discussed under point 3 in this response. However I suggest substituting the concept of "comparative advantage" for **"surprising extent".** In making comparative judgments on all four criteria categories, panel members and submitters will need to know what "average" results are, and what other programs are able to achieve as benchmarks using several indicators of success. Naturally, a panel will be more likely to judge a program as exemplary if there is evidence to substantiate several impressive positive outcome claims rather than a small or trivial claim. As panels gain experience with programs in their topic areas (and as the programs and their evidence of success improve), the panels may raise their expectations for what is exemplary.

How many programs will meet the criteria for "promising" and how many for "exemplary" remains to be discovered, as the panels get into full swing and developers become more aware of the opportunities. Based on informal discussions and early experience with two panels, I disagree with Pogrow's assessment, ***"Given the dearth of exemplary national programs, Klein's proposal to disseminate information about such programs is of little value."*** (p.22).

3. **Pogrow made important points about evidence of effectiveness that the System of Expert Panels is starting to address.** They are:

The need to be honest about negative results as well as claims of positive impact: I agree with Pogrow (p. 23) that panels are not likely to find evidence of **"a program that is universally successful"** or to have comparison information on the many implementations available, and that it is important for potential consumers to know when and in what situations the program is not likely to yield positive results. The developers of the guide for the System of Expert Panels will be asked to show how submitters can report on negative results so that they will gain credibility with the reviewers, and to discuss how to support claims of positive impact by explaining variation of results among sites.

Attention to consistency of results and creative evaluation evidence: Pogrow is correct that the panel members need to consider many aspects of evidence to judge impact claims. Pogrow's discussions of consistency and evaluation designs show that he is influenced by the same reality as the first two expert panels. In some cases, panels found a rigorous control group type study in one of the early pilot tests for the program or an occasional dissertation on a few sites, but they didn't find evaluations of nationally representative sites or comprehensive information on implementation and outcomes from the universe of replication/adaptation sites. As the System matures, expectations for more comprehensive evidence from multiple sites to support claims for exemplary programs may increase.

---

[5] Such as clinical trials or beta-testing or advanced field testing in Slavin's Design Competition proposal.

So far, panelists seem to agree with Slavin (1998) that traditional experimental designs are valuable ***but not the only way*** to demonstrate that findings can be attributed to the program, and disagree with Pogrow's statement that "***While experimental designs are essential for testing theory, classical control group designs are of little practical value for determining whether programs are exemplary.***" (p. 23). Other parts of Pogrow's discussion show he believes that while most existing evidence may not be based on gold standard classical randomized experimental designs, there are ways that it can be used to support claims of positive impact. The guidelines for the System of Expert Panels should discuss how to use an extensive mosaic of evidence to support important claims using patterns of consistent findings, logical comparisons, and appropriate analyses, etc. Panel expectations for convincing evidence showing that the claimed results can be attributed to the program will be higher for exemplary than for promising recommendations.

Need to balance panel reviews of evidence and political correctness: Pogrow, referencing a related concern from Slavin (1997, p. 21) asks how quality and significance judgments can be balanced with effectiveness judgements and not just "***reflect the current Zeitgeist about what effective practice should be.***" (p.22.) The Panels are struggling with this dilemma. They need to follow the System regulations that specify use of all four criteria categories. This means that to be designated as promising or exemplary, a program must have adequate evidence of success, while at the same time allowing judgments on what the panel experts know about research based principles in their topic areas, particularly as they look at aspects of quality and educational significance. As an example, the Mathematics and Science Expert Panel asked its first set of practitioner reviewers to examine quality, educational significance and replicability criteria. General agreement with the national mathematics standards was one subcriteria under educational significance. Other panels encourage discussion of the relation between research and program design, but do not require consistency with current beliefs or expectations on what should work. In all cases, however, evidence has to be convincing that the program does work. Datta, in a paper co-authored with Scriven (1997) agrees that panels need to guard against an over-emphasis on political correctness.[6]

---

[6] Datta says, "*The panels should be a venue for iconoclastic, original, outrageous thought on how to do it better, approaches that break the mold, that may go counter to established beliefs or challenge political correctness. The research syntheses should not become a template for screening out programs whose assumptions and approaches do not conform to current beliefs. The focus must be on results---not process. That is, the effectiveness criterion is not whether the approach uses the processes the panel thinks should work. The criterion should be results: sound evidence that the approach DOES work. Extrapolating from previous research and preconceptions about what oughtta work, should work, might work is no substitute for evidence that something actually works.* (p. 3) — She later says, "*Where national or state standards are based on first-rate state-of-the-art research, they may be sensible proxies for other information about quality*" (p. 5). But explains that standards should not be used if they are obsolete or mostly based on political consensus with only a little research. Instead of judging the quality of a program on these standards, she suggests that the System should "*Provide consumer information in a factual way about where a program does or does not conform to national standards.*"(p. 6).

4. **Pogrow says there should be more consistent Federal support for development and dissemination**. I agree, but both development and dissemination need to receive new and different types of support. Both have to be more focused on consumer-oriented impact evaluation. Key aspects of these new funding strategies include:

   - Increased support for the basic review and coordination functions of the System of Expert Panels by expanding the numbers of panels and individuals involved. To the extent possible, these constituency led panels need to work with Federal offices with funding for related research, development, dissemination and implementation in the topic area.

   - Instead of supporting large-scale demonstration projects, which primarily benefit the recipient site and are rarely tested for replication, support for development should focus on designing and evaluating models so that they can be designated as promising and exemplary by appropriate expert panels. Investments in this development should be guided by an understanding of the gaps and needs for solutions in given topic areas, much as Slavin (1997) has suggested in start-up considerations for performance-based design competitions.

   - Dissemination should be focused on consumers' needs for evaluative information. That means that the criteria and information provided on replicable promising or exemplary programs should help the potential users learn about strengths and weaknesses of these options. The System of Expert Panels plans to do this via individual consumer summary profiles of promising and exemplary programs. After two or three review cycles, panels should be able to prepare consumer reports comparing the promising and exemplary programs, to help users make choices and to identify gaps where new programs are needed.

   - Support for implementation of promising and exemplary programs should be bountiful, but tied to developer and user requirements for evaluation feedback over a multi-year period. This evaluation feedback should be particularly intensive for promising programs, so that evidence can be obtained to support a subsequent panel recommendation of exemplary. However, evaluation feedback will also be essential for exemplary programs, since developers will want to improve their claims of positive impact over the years and maintain their exemplary status as recertification requirements are developed. ED is starting to address this challenge, but this work might be accelerated as opportunities for new types of Federal support become feasible with the upcoming reauthorizations of the elementary and secondary education legislation and of OERI.

In summary, the System of Expert Panels, and its strategies to tie evaluation to dissemination and development, seem well worth trying as one reasonable way for the

Federal government to improve education. It is an intelligent way for OERI to work with its constituents within and outside the Department to identify and share promising and exemplary programs. Please share your perspectives and comments on how to make this long term strategy have positive results for all by writing to me at OERI, 555 New Jersey Ave., NW, Washington, DC 20208-5643 or e-mail: <sue_klein@ed.gov>.

References

Klein, S.S., Compiler. (Dec. 1997 a) *Disseminating Promising and Exemplary Programs: Planning a System of Expert Panels. Working Papers,* U.S. Department of Education, Office of Educational Research and Improvement, Washington, D.C.

Klein, S.S., (Aug.-Sept 1997 b) Response: A System of Expert Panels and Design Competitions: Complementary Federal Approaches to Find, Develop, and Share promising and exemplary products and programs, *Educational Researcher*, pp. 12-22.

Klein, S.S., Compiler. (May 1998) *Learning from Consumer Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs, Working Papers, Vol. 1-2,* U.S. Department of Education, Office of Educational Research and Improvement, Washington, D.C.

Pogrow, S. (Oct. 1998) What is an Exemplary Program and Why Should Anyone Care? A Reaction to Slavin and Klein, *Educational Researcher.*

Scriven, M. and Datta, L. (February 1997). Incorporating Research Findings and Practices in Expert Panel Work: A Dialog Between Michael Scriven and Lois-ellin Datta in *S. Klein, Compiler, Learning from Consumer Oriented Review Efforts to Guide the Development of a System of Expert Panels to Identify and Share Promising and Exemplary Products and Programs,* Working Papers, Vol. 1, May 1998, U.S. Department of Education, Office of Educational Research and Improvement, Washington, D.C.

Slavin, R.E., (Jan-Feb. 1997) Design Competitions: A Proposal for a New Federal Role in Educational Research and Development, *Educational Researcher*, pp. 22-28.

Slavin, R.E., (Nov. 1998?). Yes, Control Groups are Essential in Program Evaluation: A Response to Pogrow. *Educational Researcher*, pp.

U.S. Department of Education. Nov. 17, 1997 *Federal Register* Regulations for the System of Expert Panels <http://www.ed.gov/legislation/FedRegister/finrule/1997-4111797b.html>).

☆U.S. GOVERNMENT PRINTING OFFICE: 1999 - 449-481

United States
Department of Education
Washington, DC 20208–5643

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (B)

ORAD 1999-1203-II

650